

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
30 May 2003 (30.05.2003)

PCT

(10) International Publication Number  
**WO 03/044179 A2**

(51) International Patent Classification<sup>7</sup>: C12N  
(21) International Application Number: PCT/US02/37626  
(22) International Filing Date:  
20 November 2002 (20.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/332,015 20 November 2001 (20.11.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/332,015 (CIP)  
Filed on 20 November 2001 (20.11.2001)

(71) Applicant (for all designated States except US): CORVAS INTERNATIONAL, INC. [US/US]; 3030 Science Park Road, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): MADISON, Edwin, L. [US/US]; 11005 Cedarcrest Way, San Diego, CA 92121 (US). ONG, Edgar, O. [CA/US]; 10738 Glendover Lane, San Diego, CA 92126 (US).

(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 4350 La Jolla Village Drive, 7th Floor, San Diego, CA 92122-1246 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PI, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE 17, THE ENCODED POLYPEPTIDES AND METHODS BASED THEREON

(57) Abstract: Provided herein are polypeptides designated CVSP17 polypeptides that exhibit protease activity as a single chain or as an activated two chain form. Methods using the polypeptides to identify compounds that modulate the protease activity thereof are provided. The polypeptides also serve as tumor markers.

WO 03/044179 A2

-1-

**NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE 17, THE ENCODED  
POLYPEPTIDES AND METHODS BASED THEREON****RELATED APPLICATIONS**

Benefit of priority is claimed to U.S. provisional application Serial No.  
5 60/332,015, filed November 20, 2001, to Edwin L. Madison and Edgar O. Ong,  
entitled "NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE 17, THE  
ENCODED PROTEINS AND METHODS BASED THEREON." Where permitted,  
the subject matter of each of U.S. provisional application and U.S. application  
Serial No. (attorney docket no. 24745-1622PC), filed on the same day herewith,  
10 entitled "NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE 17, THE  
ENCODED POLYPEPTIDES AND METHODS BASED THEREON", is incorporated  
by reference in its entirety.

**FIELD OF THE INVENTION**

Nucleic acid molecules that encode proteases and portions thereof,  
15 particularly protease domains are provided. Also provided are prognostic,  
diagnostic and therapeutic methods using the proteases and domains thereof and  
the encoding nucleic acid molecules.

**BACKGROUND OF THE INVENTION AND OBJECTS THEREOF**

Cancer is a leading cause of death in the United States, developing in one  
20 in three Americans; one of every four Americans dies of cancer. Cancer is  
characterized by an increase in the number of abnormal neoplastic cells, which  
proliferate to form a tumor mass, the invasion of adjacent tissues by these  
neoplastic tumor cells, and the generation of malignant cells that metastasize via  
the blood or lymphatic system to regional lymph nodes and to distant sites.

25 Among the hallmarks of cancer is a breakdown in the communication  
among tumor cells and their environment. Normal cells do not divide in the  
absence of stimulatory signals, and cease dividing in the presence of inhibitory  
signals. Growth-stimulatory and growth-inhibitory signals are routinely  
exchanged between cells within a tissue. In a cancerous, or neoplastic, state, a  
30 cell acquires the ability to "override" these signals and to proliferate under  
conditions in which normal cells do not grow.

-2-

In order to proliferate tumor cells acquire a number of distinct aberrant traits reflecting genetic alterations. The genomes of certain well-studied tumors carry several different independently altered genes, including activated oncogenes and inactivated tumor suppressor genes. Each of these genetic changes appears to be responsible for imparting some of the traits that, in the aggregate, represent the full neoplastic phenotype.

A variety of biochemical factors have been associated with different phases of metastasis. Cell surface receptors for collagen, glycoproteins such as laminin, and proteoglycans, facilitate tumor cell attachment, an important step in invasion and metastases. Attachment triggers the release of degradative enzymes which facilitate the penetration of tumor cells through tissue barriers. Once the tumor cells have entered the target tissue, specific growth factors are required for further proliferation. Tumor invasion and progression involves a complex series of events, in which tumor cells detach from the primary tumor, break down the normal tissue surrounding it, and migrate into a blood or lymphatic vessel to be carried to a distant site. The breaking down of normal tissue barriers is accomplished by the elaboration of specific enzymes that degrade the proteins of the extracellular matrix that make up basement membranes and stromal components of tissues.

A class of extracellular matrix degrading enzymes has been implicated in tumor invasion. Among these are the matrix metalloproteinases (MMP). For example, the production of the matrix metalloproteinase stromelysin is associated with malignant tumors with metastatic potential (see, *e.g.*, McDonnell *et al.* (1990) *Smnrs. in Cancer Biology* 1:107-115; McDonnell *et al.* (1990) *Cancer and Metastasis Reviews* 9:309-319).

The capacity of cancer cells to metastasize and invade tissue is facilitated by degradation of the basement membrane. Several proteinase enzymes, including the MMPs, have been reported to facilitate the process of invasion of tumor cells. MMPs are reported to enhance degradation of the basement membrane, which thereby permits tumorous cells to invade tissues. For example, two major metalloproteinases having molecular weights of about 70 kDa and 92 kDa appear to enhance ability of tumor cells to metastasize.

-3-

### Serine Proteases

Serine proteases (SPs) have been implicated in neoplastic disease progression. Most serine proteases, which are either secreted enzymes or are sequestered in cytoplasmic storage organelles, have roles in blood coagulation, wound healing, digestion, immune responses and tumor invasion and metastasis.

A class of cell surface proteins designated type II transmembrane serine proteases, which are membrane-anchored proteins with additional extracellular domains, has been identified. As cell surface proteins, they are positioned to play a role in intracellular signal transduction and in mediating cell surface proteolytic events. Other serine proteases can be membrane bound and function in a similar manner. Others are secreted. Many serine proteases exert their activity upon binding to cell surface receptors, and, hence act at cell surfaces. Cell surface proteolysis is a mechanism for the generation of biologically active proteins that mediate a variety of cellular functions.

Serine proteases, including secreted and transmembrane serine proteases, have been implicated in processes involved in neoplastic development and progression. While the precise role of these proteases has not been fully elaborated, serine proteases and inhibitors thereof are involved in the control of many intra- and extracellular physiological processes, including degradative actions in cancer cell invasion, metastatic spread, and neovascularization of tumors, that are involved in tumor progression. It is believed that proteases are involved in the degradation of extracellular matrix (ECM) and contribute to tissue remodeling, and are necessary for cancer invasion and metastasis. The activity and/or expression of some proteases have been shown to correlate with tumor progression and development.

For example, a membrane-type serine protease MTSP1 (also called matriptase; see SEQ ID Nos. 1 and 2 from U.S. Patent No. 5,972,616; and GenBank Accession No. AF118224; (1999) *J. Biol. Chem.* 274:18231-18236; U.S. Patent No. 5,792,616; see, also Takeuchi (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96:11054-1161) that is expressed in epithelial cancer and normal tissue (Takeuchi *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96:11054-61) has been identified. Matriptase was originally identified in human breast cancer cells as a

-4-

major gelatinase (see, U.S. Patent No. 5,482,848), a type of matrix metalloprotease (MMP). It has been proposed that it plays a role in the metastasis of breast cancer. Matrilysin also is expressed in a variety of epithelial tissues with high levels of activity and/or expression in the human gastrointestinal tract and the prostate. MTSPs, designated MTSP3, MTSP4, MTSP6 have been described in published International PCT application No. WO 01/57194, based in International PCT application No. PCT/US01/03471.

Prostate-specific antigen (PSA), a kallikrein-like serine protease, degrades extracellular matrix glycoproteins fibronectin and laminin, and, has been postulated to facilitate invasion by prostate cancer cells (Webber *et al.* (1995) *Clin. Cancer Res.* 1:1089-94). Blocking PSA proteolytic activity with PSA-specific monoclonal antibodies results in a dose-dependent decrease *in vitro* in the invasion of the reconstituted basement membrane Matrigel by LNCaP human prostate carcinoma cells which secrete high levels of PSA.

Hepsin, a cell surface serine protease identified in hepatoma cells, is overexpressed in ovarian cancer (Tanimoto *et al.* (1997) *Cancer Res.*, 57:2884-7). The hepsin transcript appears to be abundant in carcinoma tissue and is almost never expressed in normal adult tissue, including normal ovary. It has been suggested that hepsin is frequently overexpressed in ovarian tumors and therefore can be a candidate protease in the invasive process and growth capacity of ovarian tumor cells.

A serine protease-like gene, designated normal epithelial cell-specific 1 (NES1) (Liu *et al.*, *Cancer Res.*, 56:3371-9 (1996)) has been identified. Although expression of the NES1 mRNA is observed in all normal and immortalized nontumorigenic epithelial cell lines, the majority of human breast cancer cell lines show a drastic reduction or a complete lack of its expression. The structural similarity of NES1 to polypeptides known to regulate growth factor activity and a negative correlation of NES1 expression with breast oncogenesis suggest a direct or indirect role for this protease-like gene product in the suppression of tumorigenesis.

Hence transmembrane and other serine proteases and other proteases appear to be involved in the etiology and pathogenesis of tumors. There is a

-5-

need to further elucidate their role in these processes and to identify additional transmembrane proteases. Therefore, it is an object herein to provide serine protease proteins and nucleic acids encoding such proteases, including those that are involved in the regulation of or participate in tumorigenesis and/or carcinogenesis. It is also an object herein to provide prognostic, diagnostic, therapeutic screening methods using such proteases and the nucleic acids encoding such proteases.

#### SUMMARY

Provided herein are polypeptides designated CVSP17s, including the protease domain thereof (see, *e.g.*, SEQ ID Nos. 5 and 6, particularly, for example, amino acids 105-332 of SEQ ID No. 6 or amino acids 104-332, where activation cleavage is between the R<sub>104</sub> and I<sub>105</sub>; or a variant where there is an Arg at position 258 in place of a Glu;). CVSP17 is a member of the serine protease family whose functional activity differs in tumor cells from non-tumor cells in the same tissue. CVSP17 is expressed as a secreted protein and may also bind to cell surface receptors and function as a cell-surface bound protease, such as by dimerization or multimerization with a membrane-bound or receptor-bound protein. Sequence analysis indicates the presence of a sequence of amino acids at the C-terminus that is consonant with a leucine zipper, which facilitate dimerization, and hence it may have a regulatory function as well. The CVSP17 can form homodimers and can also form heterodimers with some other protein, such as a membrane-bound protein.

CVSP17 has a signal peptide, protease domain, and a C-terminal region (amino acids 333- to 635 in an exemplified embodiment) that includes three leucine zipper (*e.g.*, aa 432-453; aa 439-460; aa 446-467 in SEQ ID No. 6 in the exemplified embodiment). Also provided are dimers and other multimers of the CVSP17 polypeptide and proteolytically active portions and fragments of the polypeptide. Single and two chain activated forms of the polypeptide are also provided as are truncated proteolytically active portions thereof, especially those that include all or of the protease domain and a portion of the C-terminal domain that contains at least a region corresponding to the region of the exemplified polypeptide that contains residues 397-427.

-6-

In particular, substantially purified single chain and two chain activated CVSP17 polypeptides and full-length CVSP17 polypeptides that include at least one leucine zipper in the C-terminus are provided. These polypeptides include a protease domain of CVSP17 or a catalytically active portion thereof, and include  
5 polypeptides such as, but are not limited to,

a) a polypeptide that contains at least 8, 10, 15, 20, 30 or more contiguous amino acids from residues 397-427 of SEQ ID No. 6 or contains 8, 10, 15, 20, 30 or more contiguous amino acids encoded by a sequence of nucleotides that hybridizes under conditions of high stringency to a sequence of  
10 nucleotides that encodes residues 397-427 of SEQ ID No. 6 (or a variant thereof where there is an Arg at position 258 in place of a Glu); or

b) the CVSP17 portion of the polypeptide consists essentially of the protease domain of the CVSP17 or a catalytically active portion thereof with the proviso that the protease domain does not include the contiguous sequence Cys  
15 Arg Ser Thr Arg Ser (SEQ ID No. 18);

c) the polypeptide contains only (consists essentially) of residues 19-332 of SEQ ID No. 6 (or a variant thereof where there is an Arg at position 258 in place of a Glu);

d) the polypeptide is encoded by a sequence of nucleotides that  
20 hybridizes along at least 70% of its full length to a sequence of nucleotides that encodes a polypeptide of any of a)-c);

e) a full-length CVSP17 polypeptide that includes a leucine zipper; or

f) the polypeptide has at least 60% sequence identity with a polypeptide of any of a)-d).

25 Included among such CVSP17 polypeptides or portions thereof with amino acid changes such that the specificity and/or protease activity remains substantially unchanged or is about 1%, 5%, 10% or more of a wild-type protein. These polypeptides include those that contain a sequence of amino acids that has at least 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%,  
30 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identity to the CVSP17 of SEQ ID No. 6 or a variant thereof where there is an Arg at position 258 in place of a Glu (or polypeptides a)-f)), where the

-7-

percentage identity is determined using standard algorithms and gap penalties that maximize the percentage identity. A human CVSP17 polypeptide is exemplified, although other mammalian CVSP17 polypeptides are contemplated. Splice variants of the CVSP17, particularly those with a proteolytically active  
5 protease domain, are contemplated herein. Dimers and other multimers of such polypeptides are also provided.

Full-length and portions of the full-length polypeptide are provided. Polypeptides that include the protease domains are provided. Such polypeptides include, but are not limited to, the single chain region having an N-terminus at  
10 the cleavage site for activation of the zymogen, through the C-terminus, or N-terminal or C-terminal truncated portions thereof that exhibit proteolytic activity as a single-chain polypeptide in *in vitro* proteolysis assays, of any family member, including CVSP17, such as from a mammal, including human, that, for example, is expressed or is active in tumor cells at different levels from non-  
15 tumor cells.

CVSP17 is expressed or activated in cervical tumors. It is also expressed in colon carcinoma tissue and pancreas islet cell tumor tissue. It may also be expressed and/or activated in other tumors, such as breast, prostate, lung, stomach, uterine, ovarian and prostate tumors and in leukemias and lymphomas.  
20 The expression and/or activation and/or secretion of the expressed protein (zymogen) of this protein can be used to monitor cancer and cancer therapy, particularly in cervical cancers. As a protease it can be involved in tumor progression. By virtue of its functional activity it can be a therapeutic or diagnostic target. The expression and/or activation (or reduction in level of  
25 expression or activation) of the expressed protein or zymogen form thereof can be used to monitor cancer and cancer therapy. For example, the expression of the this protein can be used to monitor prostate cancer and prostate cancer therapy.

The serine protease family includes members that are activated and/or  
30 expressed in tumor cells at different levels from non-tumor cells; and those from cells in which substrates therefor differ in tumor cells from non-tumor cells or in which other factors, such as co-factors, alter the specificity or activity of the

-8-

serine protease (SP). The serine protease provided herein, designated herein as CVSP17, is a secreted protease that can be membrane anchored. The protease domain and full-length protein, including the zymogen and activated forms, and uses thereof are also provided. Proteins encoded by splice variants are also  
5 provided. Nucleic acid molecules encoding the proteins and protease domains are also provided.

Nucleic acid molecules encoding the proteins and protease domains are also provided. In particular, nucleic acid molecules encoding CVSP17 from animals, including splice variants thereof are provided. The encoded proteins are  
10 also provided. Also provided are functional domains thereof. For example, the SP protease domains, portions thereof, and muteins thereof from or based on animal SPs, including, but are not limited to, rodent, such as mouse and rat; fowl, such as chicken; ruminants, such as goats, cows, deer, sheep; ovine, such as pigs; and humans. Exemplary nucleic acid encoding the CVSP17 protease  
15 and upstream nucleic acid is set forth in SEQ ID No. 5; and the encoded protein is set forth in SEQ ID No. 6 (or a variant thereof where there is an Arg at position 258 in place of a Glu). The protease domain encompasses amino acids 104-332 of SEQ ID No. 6. The serine protease histidine active site domain is amino acids 141-146 of SEQ ID No. 6 (LTAAHC). The nucleic acid and amino  
20 acid sequences of an exemplary CVSP17 are set forth in SEQ ID Nos. 5 and 6. Nucleic acid molecules that encode a single-chain protease domain or catalytically active portion thereof and also those that encode the full-length CVSP17 (SEQ ID Nos. 5 and 6) are provided. Single amino acid changes are contemplated; for example peptides in which there is an Arg at position 258 in  
25 place of a Gly are provided.

CVSP17 polypeptides, including, but not limited to those encoded by splice variants thereof, and nucleic acids encoding CVSPs, and domains, derivatives and analogs thereof are provided herein. Single chain protease domains that contain the N-termini that are generated by activation of the  
30 zymogen form of CVSP17 are also provided. The cleavage site for the protease domain in the exemplified embodiment is (R↓IVGG) (see SEQ ID Nos. 5 and 6, amino acid residues 104-332).

Nucleic acid molecules that encode a single-chain protease domain or catalytically active portion thereof and also those that encode the full-length CVSP17 are provided. Also provided are nucleic acid molecules that hybridize to such CVSP17 encoding nucleic acid along their full length or along at least about  
5 70%, 80% or 90% of the full length and encode the full length or a truncated portion thereof, such as without the signal sequence or a protease domain or catalytically active portion thereof. Hybridization is typically performed under conditions of at least low, generally at least moderate, and often high stringency.

Also provided are plasmids containing any of the nucleic acid molecules  
10 provided herein. Cells containing the plasmids are also provided. Such cells include, but are not limited to, bacterial cells, yeast cells, fungal cells, plant cells, insect cells and animal cells. In addition to cells and plasmids containing nucleic acid encoding the CVSP17 polypeptide, methods of expression of the encoded polypeptide are provided. Also provided is a method of producing CVSP17 by  
15 growing the above-described cells under conditions whereby the CVSP17 is expressed by the cells, and recovering the expressed CVSP17 polypeptide. Methods for isolating nucleic acid encoding other CVSP17s are also provided.

Also provided are cells, generally eukaryotic cells, such as mammalian cells and yeast cells, in which the CVSP17 polypeptide is expressed by the cells.  
20 Such cells to which the secreted protein can bind are used in drug screening assays to identify compounds that modulate the activity of the CVSP17 polypeptide. These assays include *in vitro* binding assays, and transcription based assays in which signal transduction mediated directly or indirectly, such as via activation of pro-growth factors, by the CVSP17 or cleavage products  
25 thereof is assessed.

The protease domain for use in the methods and assay provided herein does not have to result from activation, which produces a two chain activated product, but rather is a single chain polypeptide with an N-terminus that includes the consensus sequence ↓VVGG, ↓IVGG, ↓VGLL, ↓ILGG, ↓ITGG or ↓IVNG or  
30 other such motif at the N-terminus. Such polypeptides, although not the result of cleavage activation and not two-chain forms, exhibit proteolytic (catalytic) activity. These protease domain polypeptides, two chain and single chain forms

-10-

thereof and catalytically active fragments and longer polypeptides, are used in assays to screen for agents that modulate the activity of the CVSP17.

Such assays also are provided herein. In exemplary assays, the effects of test compounds on the ability of the full length of a single chain, two chain  
5 activated form, or a protease domain, which is a single chain or a two chain activated form, of CVSP17 to proteolytically cleave a known substrate, typically a fluorescently, chromogenically or otherwise detectably labeled substrate, are assessed. Agents, generally compounds, particularly small molecules, that modulate the activity of the protein (full length or protease domain either single  
10 or two chain forms thereof) are candidate compounds for modulating the activity of the CVSP17. The protease domains and full length proteins also can be used to produce protease-specific antibodies.

Also provided are muteins of the full-length single chain protease domain of CVSP17 particularly muteins in which the Cys residue (residue no. 211 in  
15 SEQ ID No. 6) in the protease domain that is free (*i.e.*, does not form disulfide linkages with any other Cys residue in the protease domain) is substituted with another amino acid substitution, generally with a conservative amino acid substitution or a substitution that does not eliminate the activity, and muteins in which a glycosylation site(s) is eliminated. Muteins in which other substitutions  
20 in which catalytic activity is retained are also contemplated (see, *e.g.*, Table 1, for exemplary amino acid substitutions). Generally such muteins retain at least about 1%, 2%, 3%, 5%, 7%, 8%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or more (or in increased activity, *i.e.*, 101, 102, 103, 104, 105, 110% or greater) of the protease activity of the unmutated protein.

25 Hence, provided herein is a member of the family of serine proteases designated CVSP17, and functional domains, especially protease (or catalytic) domains thereof, muteins and other derivatives and analogs thereof. Also provided herein are nucleic acids encoding the CVSP17.

Additionally provided herein are antibodies that specifically bind to the  
30 CVSP17 and inhibit the activity thereof. Included are antibodies that specifically bind to the protein or protease domain, including to the single and/or two chain forms thereof. Among the antibodies are two-chain-specific antibodies, and

-11-

single-chain specific antibodies and neutralizing antibodies. Antibodies that specifically bind to the CVSP17, particularly the single chain protease domain, the zymogen and activated form also are provided herein. Antibodies that specifically bind (*i.e.* bind with at least 2, 5 or 10-fold greater affinity compared to another protein) to the CVSP17, particularly those that specifically bind to an activated form one or both of the single-chain or two-chain forms of the protease domain or full-length two-chain form, but not to the full-length zymogen form of an CVSP17. Antibodies that specifically bind to the two-chain and/or single-chain form of CVSP17 are provided. The antibodies include those that specifically bind to the two-chain or single-chain form of the protease domain and/or the full-length protein. Also provided are antibodies that specifically bind to the leucine zipper region of a CVSP17 polypeptide.

SPs are of interest because they appear to be expressed and/or activated at different levels in tumor cells from normal cells, or have functional activity that is different in tumor cells from normal cells, such as by an alteration in a substrate therefor, or a cofactor. CVSP17 is of interest because it is expressed or is active in tumor cells. Hence the CVSP17 provided herein can serve as diagnostic markers for certain tumors. The level of activated CVSP17 can be diagnostic of uterine, pancreatic, lung, stomach, prostate or colon cancer or leukemia or other cancer.

Further provided herein are prognostic, diagnostic, and therapeutic screening methods using CVSP17 and the nucleic acids encoding CVSP17. It is shown herein, that CVSP17 is expressed in cervical cancer. It may also be expressed in colon, breast, stomach, uterine, ovarian, lung prostate tumors and in other tumors as well as in certain normal cells and tissues (see *e.g.*, EXAMPLES for tissue-specific expression profile). In particular, the prognostic, diagnostic and therapeutic screening methods are used for preventing, treating, or for finding agents useful in preventing or treating, tumors or cancers such as cervical cancer, lung carcinoma, breast cancer, colon adenocarcinoma and ovarian carcinoma.

Also provided are methods of diagnosing the presence of a pre-malignant lesion, a malignancy, or other pathologic condition in a subject, by obtaining a

-12-

biological sample from the subject; exposing it to a detectable agent that binds to a two-chain and/or single-chain form of CVSP17, where the pathological condition is characterized by the presence or absence of the two-chain or single-chain form.

5        Also provided are methods for screening for compounds that modulate the activity of CVSP17. The compounds are identified by contacting them with the CVSP17 or protease domain thereof and a substrate for the CVSP17. A change in the amount of substrate cleaved in the presence of the compounds compared to that in the absence of the compound indicates that the compound  
10       modulates the activity of the CVSP17. Such compounds are selected for further analyses or for use to modulate the activity of the CVSP17, such as inhibitors or agonists. The compounds also can be identified by contacting the substrates with a cell that binds to a CVSP17 or catalytically active portion thereof.

      Also provided herein are modulators of the activity of CVSP17, especially  
15       the modulators obtained using the screening methods provided herein. Such modulators can have use in treating cancerous conditions and other neoplastic conditions. Also provided herein are methods of modulating the activity of the CVSP17 and screening for compounds that modulate, including inhibit, antagonize, agonize or otherwise alter the activity of the CVSP17. Of particular  
20       interest is the protease domain of CVSP17 that includes the catalytic portion of the protein.

      Conjugates containing a) a CVSP17 polypeptide or protease domain in a single chain form or two chain form; and b) a targeting agent linked to the CVSP directly or via a linker, where the agent facilitates: i) affinity isolation or  
25       purification of the conjugate; ii) attachment of the conjugate to a surface; iii) detection of the conjugate; or iv) targeted delivery to a selected tissue or cell, are provided herein. The conjugate can contain a plurality of agents linked thereto. The conjugate can be a chemical conjugate; and it can be a fusion protein.

30       In another embodiment, the targeting agent is a protein or peptide fragment. The protein or peptide fragment can include a protein binding

-13-

sequence, a nucleic acid binding sequence, a lipid binding sequence, a polysaccharide binding sequence, or a metal binding sequence.

Methods of inhibiting tumor invasion or metastasis or treating a malignant or pre-malignant condition by administering an agent that inhibits activation of the zymogen form of CVSP17 or an activity of an activated form are provided. The conditions include, but are not limited to, a condition, such as a tumor, of the uterus, stomach and also the breast, cervix, prostate, esophagus, lung, ovary and colon.

Methods of diagnosing a disease or disorder characterized by detecting an aberrant level of a CVSP17 in a subject are provided. The method can be practiced by measuring the level of the DNA, RNA, protein or functional activity of the CVSP17. An increase or decrease in the level of the DNA, RNA, protein or functional activity of the CVSP, relative to the level of the DNA, RNA, protein or functional activity found in an analogous sample not having the disease or disorder (or other suitable control) is indicative of the presence of the disease or disorder in the subject.

Combinations are provided herein. A combination can include: a) an modulator, such as an inhibitor, of the activity of a CVSP17; and b) an anti-cancer treatment or agent. The CVSP inhibitor and the anti-cancer agent can be formulated in a single pharmaceutical composition or each is formulated in a separate pharmaceutical composition. The CVSP17 inhibitor can be an antibody or a fragment or binding portion thereof made against the CVSP17, such as an antibody that specifically binds to the protease domain, an inhibitor of CVSP17 production, or an inhibitor of CVSP17 membrane-localization or an inhibitor of CVSP17 activation. Other CVSP17 inhibitors include, but are not limited to, an antisense nucleic acid or double-stranded RNA (dsRNA), such as RNAi, encoding the CVSP17 or portions thereof, particularly a portion of the protease domain, a nucleic acid encoding at least a portion of a gene encoding the CVSP17 with a heterologous nucleotide sequence inserted therein such that the heterologous sequence inactivates the biological activity encoded CVSP17 or the gene encoding it. The portion of the gene encoding the CVSP17 typically flanks the heterologous sequence to promote homologous recombination with a genomic

-14-

gene encoding the CVSP17. Kits containing components of the combinations packaged optionally with instructions and additional reagents are also provided.

Also, provided are methods for treating or preventing a tumor or cancer in a mammal by administering to a mammal an effective amount of an inhibitor of a  
5 CVSP17, whereby the tumor or cancer is treated or prevented. The CVSP17 inhibitor used in the treatment or for prophylaxis is administered with a pharmaceutically acceptable carrier or excipient. The mammal treated can be a human. The treatment or prevention method can additionally include  
10 administering an anti-cancer treatment or agent simultaneously with or subsequently or before administration of the CVSP17 inhibitor.

Also provided are transgenic non-human animals bearing inactivated genes encoding the CVSP17 and bearing the genes encoding the CVSP17 or muteins thereof under non-native or native promotor control. Such animals are useful in animal models of tumor initiation, growth and/or progression models.  
15 For example, also provided is a recombinant non-human animal in which an endogenous gene of a CVSP17 has been deleted or inactivated by homologous recombination or other recombination events or insertional mutagenesis of the animal or an ancestor thereof. A recombinant non-human animal is provided herein, where the gene of a CVSP17 is under control of a promoter that is not  
20 the native promoter of the gene or that is not the native promoter of the gene in the non-human animal or where the nucleic acid encoding the CVSP17 is heterologous to the non-human animal and the promoter is the native or a non-native promoter or the CVSP17 is on an extrachromosomal element, such as a plasmid or artificial chromosome. Transgenic non-human animals bearing the  
25 genes encoding the CVSP17 and bearing inactivated genes encoding CVSP17, particularly under a non-native promotor control or on an exogenous element, such as a plasmid or artificial chromosome, are additionally provided herein.

Pharmaceutical compositions containing the protease domain and/or full-length or other domain of a CVSP17 polypeptide are provided herein in a  
30 pharmaceutically acceptable carrier or excipient are provided herein.

Also provided are articles of manufacture that contain CVSP17 polypeptide and protease domains of CVSP17 in single chain forms or activated

-15-

forms. Articles containing a) packaging material; b) the polypeptide (or encoding nucleic acid), particularly the single chain protease domain thereof; and c) a label indicating that the article is for using in assays for identifying modulators of the activities of a CVSP17 polypeptide, are provided herein.

5           Also provided are methods of treatments of tumors by administering a prodrug that is activated by CVSP17 that is expressed or active in tumor cells, particularly those in which its functional activity in tumor cells is greater than in non-tumor cells. The prodrug is administered and, upon administration, active CVSP17 cleaves the prodrug and releases active drug in the vicinity of the tumor  
10 cells. The active anti-cancer drug accumulates in the vicinity of the tumor. This is particularly useful in instances in which CVSP17 is expressed or active in greater quantity, higher level or predominantly in tumor cells compared with other cells.

          Also provided are methods of identifying a compound that binds to the  
15 single-chain and/or two-chain form of CVSP17, by contacting a test compound with one or both forms; determining to which form the compound binds; and if it binds to a form of CVSP17, further determining whether the compound has at least one of the following properties:

- 20           (i) inhibits activation of the single-chain zymogen form of CVSP17;
- (ii) inhibits activity of the two-chain and/or single-chain form; and
- (iii) inhibits dimerization of the protein.

The forms can be full length or truncated forms, including but not limited to, the protease domain resulting from cleavage at the RI activation site or from expression of the protease domain or catalytically active portions thereof.

25           Methods for monitoring tumor progression and/or therapeutic effectiveness are also provided. The levels of activation or expression or activity of a CVSP17 polypeptide or a protease domain thereof are assessed, and the change in one or more of these levels, reflects tumor progression and/or the effectiveness of therapy. Generally, as the tumor progresses the amount of  
30 CVSP17 in a body tissue or fluid sample increases; effective therapy reduces the level.

## DETAILED DESCRIPTION OF THE INVENTION

### A. DEFINITIONS

- Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference in their entirety. In the event that there are a plurality of definitions for terms herein, those in this section prevail.
- Where reference is made to a URL or other such identifier or address, it is understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.
- As used herein, the abbreviations for any protective groups, amino acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) *Biochem.* 11:942-944).
- As used herein, serine protease refers to a diverse family of proteases wherein a serine residue is involved in the hydrolysis of proteins or peptides. The serine residue can be part of the catalytic triad mechanism, which includes a serine, a histidine and an aspartic acid in the catalysis, or be part of the hydroxyl/ $\epsilon$ -amine or hydroxyl/ $\alpha$ -amine catalytic dyad mechanism, which involves a serine and a lysine in the catalysis. Of particular interest are SPs of mammalian, including human, origin. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*, Watson *et al.* (1987) *Molecular Biology of the Gene*, 4th Edition, The Benjamin/Cummings Pub. co., p.224).
- As used herein, "transmembrane serine protease (MTSP)" refers to a family of transmembrane serine proteases that share common structural features as described herein (see, also Hooper *et al.* (2001) *J. Biol. Chem.* 276:857-860).

-17-

Thus, reference, for example, to "MTSP" encompasses all proteins encoded by the MTSP gene family, including but are not limited to: MTSP3, MTSP4, MTSP6, MTSP7, MTSP9, MTSP10, MTSP20 or an equivalent molecule obtained from any other source or that has been prepared synthetically or that exhibits the same activity. Other MTSPs include, but are not limited to, corin, enteropeptidase, human airway trypsin-like protease (HAT), MTSP1, TMPRSS2 and TMPRSS4. Sequences of encoding nucleic acid molecules and the encoded amino acid sequences of exemplary MTSPs and/or domains thereof are set forth, for example in U.S. application Serial No. 09/776,191 (SEQ ID Nos. 1-12, 49, 50 and 61-72 therein, published as International PCT application No. WO 01/57194; see also published International PCT application Nos. WO 02/072786 and WO 02/977267, and International PCT application Nos. PCT/US02/21208 and PCT/US02/15332). The term also encompass MTSPs with amino acid substitutions that do not substantially alter activity of each member and also encompasses splice variants thereof. Suitable substitutions, including, although not necessarily, conservative substitutions of amino acids, are known to those of skill in this art and can be made without eliminating a biological activity, such as the catalytic activity, of the resulting molecule.

As used herein, Type I MTSP refers to transmembrane proteins made with an N-terminal signal peptide that is cleaved so that the new N-terminus is on the extracytoplasmic side of the membrane. The original N-terminus likely stays on the cytoplasmic side, and cleavage occurs on the other side of the membrane. These proteins are anchored through a C-terminal membrane-spanning segment.

As used herein, Type II MTSP refers to transmembrane proteins that are synthesized with N-terminal or internal signal peptides that are not cleaved and that serve as a membrane anchor.

As used herein, a "protease domain of a CVSP", particularly CVSP17, refers to a domain of an SP that exhibits proteolytic activity and shares homology and structural features with the chymotrypsin/trypsin family protease domains. Hence it is at least the minimal portion of the domain that exhibits proteolytic activity as assessed by standard *in vitro* assays. Those of skill in this art recognize that a protease domain is the portion of the protease that is

-18-

structurally equivalent to the trypsin or chymotrypsin fold. Contemplated herein are polypeptides that include such protease domains and catalytically active portions thereof. Also provided are truncated forms of the protease domain that include the smallest fragment thereof that acts catalytically as a single chain  
5 form.

As used herein, the catalytically active domain of a CVSP refers to the protease domain. Reference to the protease domain of a CVSP refers to the single chain form of the protein. If the two-chain form or both is intended, it is so-specified. The zymogen form of each protein is a single chain, which is  
10 converted to the active two chain form by activation cleavage.

As used herein a CVSP17, whenever referenced herein, includes at least one or all of or any combination of:

a polypeptide encoded by the sequence of nucleotides set forth in SEQ ID No. 5 (or a variant thereof that encodes an Arg at position 258 in place  
15 of a Glu);

a polypeptide encoded by a sequence of nucleotides that hybridizes under conditions of low, moderate or high stringency to the sequence of nucleotides set forth in SEQ ID No. 5 (or a variant thereof that encodes an Arg at position 258 in place of a Glu);

20 a polypeptide that comprises the sequence of amino acids set forth in SEQ ID No. 6 (or a variant thereof where there is an Arg at position 258 in place of a Glu);

a polypeptide that comprises a sequence of amino acids having at least about 60%, 70%, 80%, 90% or about 95% sequence identity with the  
25 sequence of amino acids set forth in SEQ ID No. 6 (or a variant thereof where there is an Arg at position 258 in place of a Glu); and/or

a polypeptide encoded by a splice variant of a sequence of nucleotides that encodes a CVSP17.

By reference to SEQ ID No. 6, it is understood that a variant CVSP17 in  
30 which there is an Arg at position 258 in place of a Glu is also provided.

-19-

In particular, CVSP17 polypeptides, as provided herein are those that include a protease domain of serine protease 17 (CVSP17) or a catalytically active portion thereof, where:

- a) the polypeptide also includes at least 10 or more contiguous amino acids from residues 397-427 of SEQ ID No. 6 or comprises 10 or more contiguous amino acids encoded by a sequence of nucleotides that hybridizes under conditions of high stringency to a sequence of nucleotides that encodes residues 397-427 of SEQ ID No. 6; or
- b) the CVSP17 portion of the polypeptide is only the protease domain of a CVSP17 or a catalytically active portion thereof, except that the protease domain does not include Cys Arg Ser Thr Arg Ser (SEQ ID No. 18) as a contiguous sequence;
- c) the polypeptide contains only residues 19-332 of SEQ ID No. 6;
- d) the polypeptide contains the sequence of amino acids set forth in SEQ ID No. 6;
- e) the polypeptide is encoded by a sequence of nucleotides that hybridizes under conditions of at least moderate, and can be high, stringency along at least 70% of its full length to a sequence of nucleotides than encodes a polypeptide of any of a)-e); and/or
- f) the polypeptide has at least 60%, 60%, 70%, 80%, 90% or about 95% sequence identity with the sequence identity with a polypeptide of any of a)-e). Smaller portions thereof that retain protease activity are contemplated.

The CVSP17 can be from any animal, particularly a mammal, and includes but are not limited to, primates including humans, gorillas and monkeys; rodents, such as mice and rats; fowl, such as chickens; ruminants, such as goats, cows, deer, sheep; ovine, such as pigs and other animals. The full length zymogen or two-chain activated form is contemplated or any domain thereof, including the protease domain, which can be a two-chain activated form, or a single chain form. An exemplary CVSP17 protein includes the sequence of amino acids set forth in SEQ ID No. 6; the protease domain is set forth as amino acids 105-332 in SEQ ID No. 6.

-20-

As used herein a protease domain of a CVSP17, whenever referenced herein, includes at least one or all of or any combination of or a catalytically active portion of a CVSP17 polypeptide as defined herein. Protease domains of CVSPs vary in size and constitution, including insertions and deletions in surface loops. They retain conserved structure, including at least one of the active site triad, primary specificity pocket, oxyanion hole and/or other features of serine protease domains of proteases. Thus, for purposes herein, the protease domain is a portion of a CVSP, as defined herein, and is homologous to a domain of other SPs. As with the larger class of enzymes of the chymotrypsin (S1) fold (see, *e.g.*, Internet accessible MEROPS data base), the CVSPs protease domains share a high degree of amino acid sequence identity. The His, Asp and Ser residues necessary for activity are present in conserved motifs. The activation site, whose cleavage creates the N-terminus of protease domain in the two-chain forms has a conserved motif and readily can be identified. An exemplary protease domain of a CVSP17 is set forth as amino acids 104-332 in SEQ ID No. 6 (where the activation cleavage results in a polypeptide that contains amino acid 105-332 and beyond up to the C-terminus).

As used herein, by active form is meant a form active *in vivo* and/or *in vitro*. Single chain forms of the SPs and the catalytic domains or proteolytically active portions thereof (typically C-terminal truncations) exhibit protease activity. For example, a polypeptide containing the protease domain can exist as an activated two-chain or a single chain active form. The active single chain and two chain forms of a CVSP17 and catalytic domains or proteolytically active portions thereof can exhibit protease activity. Among the polypeptides provided herein, are isolated single chain forms and two chain forms of CVSP17 polypeptides that include protease domains and their use, for example, in *in vitro* drug screening assays for identification of agents that modulate the activity thereof.

As used herein, activation cleavage refers to the cleavage of the protease at the N-terminus of the protease domain (generally between an R and I). By virtue of the Cys-Cys pairing between a Cys outside the protease domain and a Cys in the protease domain (in the exemplified embodiment Cys<sub>88</sub> and Cys<sub>211</sub>

-21-

SEQ ID No. 6) upon cleavage the resulting polypeptide has two chains "A" chain, in this instance a chain, which in the exemplified embodiment includes at least residues 88 to 104 (or a shortened form thereof) and a "B" chain, which in the exemplified embodiment includes residues 105-211, which is the protease domain). Cleavage can be effected by another protease or autocatalytically.

As used herein, a two-chain form of the protease domain refers to a two-chain form that is formed from the a one-chain form of the protease in which the Cys pairing between a Cys outside the protease domain (*i.e.* Cys<sub>88</sub> (SEQ ID No. 6)), which links the protease domain to the remainder of the polypeptide. Upon activation cleavage, two chains are produced. For example a two chain form of a CVSP17 includes from Cys<sub>88</sub> up to and including Cys<sub>211</sub> (or beyond) of SEQ ID No. 6 where the A chain includes at least Cys<sub>88</sub> and can include up to R<sub>104</sub> and the B chain includes I<sub>105</sub> to at least Cys<sub>211</sub> and can include up to the C-terminus.

Hence provided herein are isolated single chain forms of the protease domains of SPs and their use in *in vitro* drug screening assays for identification of agents that modulate the activity thereof.

As used herein, a human protein is one encoded by nucleic acid, such as DNA, present in the genome of a human, including all allelic variants and conservative variations as long as they are not variants found in other mammals.

As used herein, a "nucleic acid encoding a protease domain or catalytically active portion of a SP" refers to a nucleic acid encoding only the recited single chain protease domain or active portion thereof, and not the other contiguous portions of the SP as a continuous sequence.

As used herein, catalytic activity refers to the activity of the SP as a serine protease. Function of the SP refers to its function in tumor biology, including promotion of or involvement in initiation, growth or progression of tumors, and also roles in signal transduction. Catalytic activity refers to the activity of the SP as a protease as assessed in *in vitro* proteolytic assays that detect proteolysis of a selected substrate.

As used herein, a leucine zipper refers to short alpha-helical coiled-coils that can form homo- and heteroligomers, such as dimers, of proteins. For example, eukaryotic transcription factors use leucine zippers to dimerize.

-22-

Accordingly, dimerized and higher multimers of CVSP17 polypeptides and/or portions thereof are provided.

As used herein, a zymogen is an inactive precursor of a proteolytic enzyme. Such precursors are generally larger, although not necessarily larger  
5 than the active form. With reference to serine proteases, zymogens are converted to active enzymes by specific cleavage, including catalytic and autocatalytic cleavage, or by binding of an activating co-factor, which generates an active enzyme. A zymogen, thus, is an enzymatically inactive protein that is converted to a proteolytic enzyme by the action of an activator.

10 As used herein, "disease or disorder" refers to a pathological condition in an organism resulting from, *e.g.*, infection or genetic defect, and characterized by identifiable symptoms.

As used herein, neoplasm (neoplasia) refers to abnormal new growth, and thus means the same as *tumor*, which can be benign or malignant. Unlike  
15 *hyperplasia*, neoplastic proliferation persists even in the absence of the original stimulus.

As used herein, neoplastic disease refers to any disorder involving cancer, including tumor development, growth, metastasis and progression.

As used herein, cancer is a general term for diseases caused by or  
20 characterized by any type of malignant tumor.

As used herein, malignant, as applies to tumors, refers to primary tumors that have the capacity of *metastasis* with loss of *growth control* and *positional control*.

As used herein, an anti-cancer agent (used interchangeable with "anti-  
25 tumor or anti-neoplastic agent") refers to any agents used in the anti-cancer treatment. These include any agents, when used alone or in combination with other compounds, that can alleviate, reduce, ameliorate, prevent, or place or maintain in a state of remission of clinical symptoms or diagnostic markers associated with neoplastic disease, tumors and cancer, and can be used in  
30 methods, combinations and compositions provided herein. Non-limiting examples of anti-neoplastic agents include anti-angiogenic agents, alkylating agents, antimetabolite, certain natural products, platinum coordination

-23-

complexes, anthracenediones, substituted ureas, methylhydrazine derivatives, adrenocortical suppressants, certain hormones, antagonists and anti-cancer polysaccharides.

As used herein, a splice variant refers to a variant produced by differential  
5 processing of a primary transcript of genomic nucleic acid, such as DNA, that results in more than one type of mRNA. Splice variants of SPs are provided herein.

As used herein, angiogenesis is intended to broadly encompass the totality of processes directly or indirectly involved in the establishment and  
10 maintenance of new vasculature (neovascularization), including, but not limited to, neovascularization associated with tumors.

As used herein, anti-angiogenic treatment or agent refers to any therapeutic regimen and compound, when used alone or in combination with other treatment or compounds, that can alleviate, reduce, ameliorate, prevent, or  
15 place or maintain in a state of remission of clinical symptoms or diagnostic markers associated with undesired and/or uncontrolled angiogenesis. Thus, for purposes herein an anti-angiogenic agent refers to an agent that inhibits the establishment or maintenance of vasculature. Such agents include, but are not limited to, anti-tumor agents, and agents for treatments of other disorders  
20 associated with undesirable angiogenesis, such as diabetic retinopathies, restenosis, hyperproliferative disorders and others.

As used herein, non-anti-angiogenic anti-tumor agents refer to anti-tumor agents that do not act primarily by inhibiting angiogenesis.

As used herein, pro-angiogenic agents are agents that promote the  
25 establishment or maintenance of the vasculature. Such agents include agents for treating cardiovascular disorders, including heart attacks and strokes.

As used herein, undesired and/or uncontrolled angiogenesis refers to pathological angiogenesis wherein the influence of angiogenesis stimulators outweighs the influence of angiogenesis inhibitors. As used herein, deficient  
30 angiogenesis refers to pathological angiogenesis associated with disorders where there is a defect in normal angiogenesis resulting in aberrant angiogenesis or an absence or substantial reduction in angiogenesis.

-24-

As used herein, the protease domain of an SP protein refers to the protease domain of an SP that exhibits proteolytic activity. Hence it is at least the minimal portion of the protein that exhibits proteolytic activity as assessed by standard assays *in vitro*. It refers to single chain forms and also to two chain  
5 activated forms (where a two chain form is intended it will be so-noted). Exemplary protease domains include at least a sufficient portion of sequences of amino acids set forth in SEQ ID No. 6 (encoded by nucleotides in SEQ ID No. 5) to exhibit protease activity.

Also contemplated are nucleic acid molecules that encode a polypeptide  
10 that has proteolytic activity in an *in vitro* proteolysis assay and that have at least 60%, 70%, 80%, 90% or about 95% sequence identity with the full length of a protease domain of a CVSP17 polypeptide, or that hybridize along their full length or along at least about 70%, 80% or 90% of the full length to a nucleic acids that encode a protease domain, particularly under conditions of moderate,  
15 generally high, stringency.

For the protease domains, residues at the N-terminus can be critical for activity. The protease domain of the single chain form of the CVSP17 protease is catalytically active. Hence the protease domain generally requires the N-terminal amino acids thereof for activity; the C-terminus portion can be  
20 truncated. The amount that can be removed can be determined empirically by testing the polypeptide for protease activity in an *in vitro* assay that assesses catalytic cleavage.

Thus, for purposes herein, the protease domain is a single chain portion of a CVSP17, as defined herein, but is homologous in its structural features and retention of sequence of similarity or homology to the protease domain of  
25 chymotrypsin or trypsin. The polypeptide exhibits proteolytic activity as a single chain.

As used herein, by homologous means about greater than 25% nucleic acid sequence identity, such as 25%, 40%, 60%, 70%, 80%, 90% or 95%. If  
30 necessary the percentage homology will be specified. The terms "homology" and "identity" are often used interchangeably. In general, sequences are aligned so that the highest order match is obtained (see, *e.g.*: *Computational Molecular*

-25-

*Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988;  
*Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence*  
5 *Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; Carillo *et al.* (1988) *SIAM J Applied Math* 48:1073).  
By sequence identity, the number of identical amino acids is determined by standard alignment algorithms programs, and are used with default gap penalties  
10 established by each supplier. Substantially homologous nucleic acid molecules would hybridize typically at moderate stringency or at high stringency all along the length of the nucleic acid or along at least about 70%, 80% or 90% of the full length nucleic acid molecule of interest. Also contemplated are nucleic acid molecules that contain degenerate codons in place of codons in the hybridizing  
15 nucleic acid molecule.

Whether any two nucleic acid molecules have nucleotide sequences that are at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson *et al.* (1988) *Proc. Natl.*  
20 *Acad. Sci. USA* 85:2444 (other programs include the GCG program package (Devereux, J., *et al.*, *Nucleic Acids Research* 12(II):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al.*, *J Molec Biol* 215:403 (1990); Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo *et al.* (1988) *SIAM J Applied Math* 48:1073). For example, the BLAST  
25 function of the National Center for Biotechnology Information database can be used to determine identity. Other commercially or publicly available programs include, DNASTar "MegAlign" program (Madison, WI) and the University of Wisconsin Genetics Computer Group (UWG) "Gap" program (Madison WI)).  
Percent homology or identity of proteins and/or nucleic acid molecules can be  
30 determined, for example, by comparing sequence information using a GAP computer program (*e.g.*, Needleman *et al.* (1970) *J. Mol. Biol.* 48:443, as revised by Smith and Waterman ((1981) *Adv. Appl. Math.* 2:482). Briefly, the

-26-

GAP program defines similarity as the number of aligned symbols (i.e., nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. Default parameters for the GAP program can include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and the weighted comparison matrix of Gribkov *et al.* (1986) *Nucl. Acids Res.* 14:6745, as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps. Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide.

As used herein, the term "at least 90% identical to" refers to percent identities from 90 to 100% relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of 100 amino acids are compared, no more than 10% (i.e., 10 out of 100) of amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons can be made between a test and reference polynucleotides. Such differences can be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they can be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions. At the level of homologies or identities above about 85-90%, the result should be independent of the program and gap parameters set; such high levels of identity can be assessed readily, often without relying on software.

As used herein, primer refers to an oligonucleotide containing two or more deoxyribonucleotides or ribonucleotides, typically more than three, from which synthesis of a primer extension product can be initiated. Experimental conditions conducive to synthesis include the presence of nucleoside triphosphates and an agent for polymerization and extension, such as DNA polymerase, and a suitable buffer, temperature and pH.

-27-

As used herein, animal includes any animal, such as, but are not limited primates including humans, gorillas and monkeys; rodents, such as mice and rats; fowl, such as chickens; ruminants, such as goats, cows, deer, sheep; ovine, such as pigs and other animals. Non-human animals exclude humans as  
5 the contemplated animal. The SPs provided herein are from any source, animal, plant, prokaryotic and fungal. Most CVSP17s are of animal origin, including mammalian origin.

As used herein, genetic therapy involves the transfer of heterologous nucleic acid, such as DNA, into certain cells, target cells, of a mammal,  
10 particularly a human, with a disorder or conditions for which such therapy is sought. The nucleic acid, such as DNA, is introduced into the selected target cells in a manner such that the heterologous nucleic acid, such as DNA, is expressed and a therapeutic product encoded thereby is produced. Alternatively, the heterologous nucleic acid, such as DNA, can in some manner  
15 mediate expression of DNA that encodes the therapeutic product, or it can encode a product, such as a peptide or RNA that in some manner mediates, directly or indirectly, expression of a therapeutic product. Genetic therapy can also be used to deliver nucleic acid encoding a gene product that replaces a defective gene or supplements a gene product produced by the mammal or the  
20 cell in which it is introduced. The introduced nucleic acid can encode a therapeutic compound, such as a growth factor inhibitor thereof, or a tumor necrosis factor or inhibitor thereof, such as a receptor therefor, that is not normally produced in the mammalian host or that is not produced in therapeutically effective amounts or at a therapeutically useful time. The  
25 heterologous nucleic acid, such as DNA, encoding the therapeutic product can be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof. Genetic therapy can also involve delivery of an inhibitor or repressor or other modulator of gene expression.

30 As used herein, heterologous nucleic acid is nucleic acid that (if DNA encodes RNA) and proteins that are not normally produced *in vivo* by the cell in which it is expressed or that mediates or encodes mediators that alter expression

-28-

of endogenous nucleic acid, such as DNA, by affecting transcription, translation, or other regulatable biochemical processes. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Heterologous nucleic acid  
5 can be endogenous, but is nucleic acid that expressed from a different locus or altered in its expression. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not normally produced by the cell or in the same way in the cell in which it is expressed. Heterologous nucleic acid, such as DNA, can also be referred to as foreign nucleic acid, such as DNA. Thus,  
10 heterologous nucleic acid or foreign nucleic acid includes a nucleic acid molecule not present in the exact orientation or position as the counterpart nucleic acid molecule, such as DNA, found in the genome. It can also refer to a nucleic acid molecule from another organism or species (*i.e.*, exogenous).

Any nucleic acid, such as DNA, that one of skill in the art would  
15 recognize or consider as heterologous or foreign to the cell in which the nucleic acid is expressed is herein encompassed by heterologous nucleic acid; heterologous nucleic acid includes exogenously added nucleic acid that is also expressed endogenously. Examples of heterologous nucleic acid include, but are not limited to, nucleic acid that encodes traceable marker proteins, such as a  
20 protein that confers drug resistance, nucleic acid that encodes therapeutically effective substances, such as anti-cancer agents, enzymes and hormones, and nucleic acid, such as DNA, that encodes other types of proteins, such as antibodies. Antibodies that are encoded by heterologous nucleic acid can be secreted or expressed on the surface of the cell in which the heterologous  
25 nucleic acid has been introduced.

As used herein, a therapeutically effective product for gene therapy is a product that is encoded by heterologous nucleic acid, typically DNA, that, upon introduction of the nucleic acid into a host, a product is expressed that ameliorates or eliminates the symptoms, manifestations of an inherited or  
30 acquired disease or that cures the disease. Also included are biologically active nucleic acid molecules, such as RNAi and antisense.

As used herein, gene therapy refers to therapy effected by the administration of a nucleic acid to a subject.

As used herein, recitation that a polypeptide consists essentially of the protease domain means that the only SP portion of the polypeptide is a protease domain or a catalytically active portion thereof. The polypeptide can optionally, and generally will, include additional non-SP-derived sequences of amino acids.

As used herein, cancer or tumor treatment or agent refers to any therapeutic regimen and/or compound that, when used alone or in combination with other treatments or compounds, can alleviate, reduce, ameliorate, prevent, or place or maintain in a state of remission of clinical symptoms or diagnostic markers associated with deficient angiogenesis.

As used herein, domain refers to a portion of a molecule, *e.g.*, proteins or the encoding nucleic acids, that is structurally and/or functionally distinct from other portions of the molecule.

As used herein, protease refers to an enzyme catalyzing hydrolysis of proteins or peptides. It includes zymogen forms and activated single and two chain forms thereof. For clarity reference to protease refers to all forms, and particular forms will be specifically designated.

As used herein, nucleic acids include DNA, RNA and analogs thereof, including peptide nucleic acids (PNA) and mixtures thereof. Nucleic acids can be single or double-stranded. When referring to probes or primers, which are optionally labeled, such as with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that their target is statistically unique or of low copy number (typically less than 5, generally less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous of sequence complementary to or identical a gene of interest. Probes and primers can be 10, 20, 30, 50, 100 or more nucleic acids long.

As used herein, a probe or primer based on a nucleotide sequence disclosed herein, includes at least 10, 14, typically at least 16 contiguous sequence of nucleotides of SEQ ID No. 5, and probes of at least 30, 50 or 100 contiguous sequence of nucleotides of SEQ ID No. 5. The length of the probe or

-30-

primer for unique hybridization is a function of the complexity of the genome of interest.

As used herein, nucleic acid encoding a fragment or portion of an SP refers to a nucleic acid encoding only the recited fragment or portion of SP, and  
5 not the other contiguous portions of the SP.

As used herein, operative linkage of heterologous nucleic to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences refers to the relationship between such nucleic acid, such as DNA, and such sequences of  
10 nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA. Thus, operatively linked or operationally associated refers to the functional  
15 relationship of nucleic acid, such as DNA, with regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences. For example, operative linkage of DNA to a promoter refers to the physical and functional relationship between the DNA and the promoter such that the transcription of such DNA is  
20 initiated from the promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA. In order to optimize expression and/or *in vitro* transcription, it can be necessary to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potentially inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that can interfere with or reduce  
25 expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak *J. Biol. Chem.* 266:19867-19870 (1991)) can be inserted immediately 5' of the start codon and can enhance expression. The desirability of (or need for) such modification can be empirically determined.

30 As used herein, a sequence complementary to at least a portion of an RNA, with reference to antisense oligonucleotides, means a sequence having sufficient complementarity to be able to hybridize with the RNA, generally under

-31-

moderate or high stringency conditions, forming a stable duplex; in the case of double-stranded SP antisense nucleic acids, a single strand of the duplex DNA (or dsRNA) can thus be tested, or triplex formation can be assayed. The ability to hybridize depends on the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with a SP encoding RNA it can contain and still form a stable duplex (or triplex, as the case can be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

10 For purposes herein, amino acid substitutions can be made in any of SPs and protease domains thereof provided that the resulting protein exhibits protease activity. Muteins can be made by making conservative amino acid substitutions and also non-conservative amino acid substitutions. For example, amino acid substitutions that desirably alter properties of the proteins can be made. In one embodiment, mutations that prevent degradation of the polypeptide can be made. Many proteases cleave after basic residues, such as R and K; to eliminate such cleavage, the basic residue is replaced with a non-basic residue. Interaction of the protease with an inhibitor can be blocked while retaining catalytic activity by effecting a non-conservative change at the site interaction of the inhibitor with the protease. Receptor binding can be altered without altering catalytic activity.

Amino acid substitutions contemplated include conservative substitutions, such as those set forth in Table 1, which do not eliminate proteolytic activity. As described herein, substitutions that alter properties of the proteins, such as removal of cleavage sites and other such sites are also contemplated; such substitutions are generally non-conservative, but can be readily effected by those of skill in the art.

Suitable conservative substitutions of amino acids are known to those of skill in this art and can be made generally without altering the biological activity, for example enzymatic activity, of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*,

-32-

Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Bejacmin/Cummings Pub. co., p.224). Also included within the definition, is the catalytically active fragment of an SP, particularly a single chain protease portion. Conservative amino acid substitutions are made, for example, in accordance with those set forth in TABLE 1 as follows:

TABLE 1

	Original residue	Conservative substitution
	Ala (A)	Gly; Ser, Abu
	Arg (R)	Lys, orn
10	Asn (N)	Gln; His
	Cys (C)	Ser
	Gln (Q)	Asn
	Glu (E)	Asp
	Gly (G)	Ala; Pro
15	His (H)	Asn; Gln
	Ile (I)	Leu; Val; Met; Nle; Nva
	Leu (L)	Ile; Val; Met; Nle; Nv
	Lys (K)	Arg; Gln; Glu
	Met (M)	Leu; Tyr; Ile; NLe Val
20	Ornithine	Lys; Arg
	Phe (F)	Met; Leu; Tyr
	Ser (S)	Thr
	Thr (T)	Ser
	Trp (W)	Tyr
25	Tyr (Y)	Trp; Phe
	Val (V)	Ile; Leu; Met; Nle; Nv

Other substitutions are also permissible and can be determined empirically or in accord with known conservative substitutions.

As used herein, Abu is 2-aminobutyric acid; Orn is ornithine.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which occur in the various DNA fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, amelioration of the symptoms of a particular disorder by administration of a particular pharmaceutical composition refers to any lessening, whether permanent or temporary, lasting or transient that can be attributed to or associated with administration of the composition.

-33-

As used herein, antisense polynucleotides refer to synthetic sequences of nucleotide bases complementary to mRNA or the sense strand of double-stranded DNA. Admixture of sense and antisense polynucleotides under appropriate conditions leads to the binding of the two molecules, or

5 hybridization. When these polynucleotides bind to (hybridize with) mRNA, inhibition of protein synthesis (translation) occurs. When these polynucleotides bind to double-stranded DNA, inhibition of RNA synthesis (transcription) occurs. The resulting inhibition of translation and/or transcription leads to an inhibition of the synthesis of the protein encoded by the sense strand. Antisense nucleic

10 acid molecule typically contain a sufficient number of nucleotides to specifically bind to a target nucleic acid, generally at least 5 contiguous nucleotides, often at least 14 or 16 or 30 contiguous nucleotides or modified nucleotides complementary to the coding portion of a nucleic acid molecule that encodes a gene of interest, for example, nucleic acid encoding a single chain protease

15 domain of an SP.

As used herein, an array refers to a collection of elements, such as antibodies, containing two or more members. An addressable array is one in which the members of the array are identifiable, typically by position on a solid phase support. Hence, in general the members of the array are immobilized on

20 discrete identifiable loci on the surface of a solid phase.

As used herein, antibody refers to an immunoglobulin, whether natural or partially or wholly synthetically produced, including any derivative thereof that retains the specific binding ability the antibody. Hence antibody includes any protein having a binding domain that is homologous or substantially homologous

25 to an immunoglobulin binding domain. Antibodies include members of any immunoglobulin claims, including IgG, IgM, IgA, IgD and IgE.

As used herein, antibody fragment refers to any derivative of an antibody that is less than full length, retaining at least a portion of the full-length antibody's specific binding ability. Examples of antibody fragments include, but

30 are not limited to, Fab, Fab', F(ab)<sub>2</sub>, single-chain Fvs (scFV), FV, dsFV diabody and Fd fragments. The fragment can include multiple chains linked together,

-34-

such as by disulfide bridges. An antibody fragment generally contains at least about 50 amino acids and typically at least 200 amino acids.

As used herein, a Fv antibody fragment is composed of one variable heavy chain domain ( $V_H$ ) and one variable light chain domain linked by noncovalent interactions.

As used herein, a dsFV refers to an Fv with an engineered intermolecular disulfide bond, which stabilizes the  $V_H$ - $V_L$  pair.

As used herein, a  $F(ab)_2$  fragment is an antibody fragment that results from digestion of an immunoglobulin with pepsin at pH 4.0-4.5; it can be recombinantly produced to produce the equivalent fragment.

As used herein, Fab fragments are antibody fragments that result from digestion of an immunoglobulin with papain; it can be recombinantly produced to produce the equivalent fragment.

As used herein, scFVs refer to antibody fragments that contain a variable light chain ( $V_L$ ) and variable heavy chain ( $V_H$ ) covalently connected by a polypeptide linker in any order. The linker is of a length such that the two variable domains are bridged without substantial interference. Included linkers are  $(Gly-Ser)_n$  residues with some Glu or Lys residues dispersed throughout to increase solubility.

As used herein, humanized antibodies refer to antibodies that are modified to include human sequences of amino acids so that administration to a human does not provoke an immune response. Methods for preparation of such antibodies are known. For example, to produce such antibodies, the encoding nucleic acid in the hybridoma or other prokaryotic or eukaryotic cell, such as an *E. coli* or a CHO cell, that expresses the monoclonal antibody is altered by recombinant nucleic acid techniques to express an antibody in which the amino acid composition of the non-variable region is based on human antibodies. Computer programs have been designed to identify such non-variable regions.

As used herein, diabodies are dimeric scFV; diabodies typically have shorter peptide linkers than scFvs, and they generally dimerize.

-35-

As used herein, production by recombinant means by using recombinant DNA methods means the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA.

As used herein the term assessing is intended to include

5 quantitative and qualitative determination in the sense of obtaining an absolute value for the activity of an SP, or a domain thereof, present in the sample, and also of obtaining an index, ratio, percentage, visual or other value indicative of the level of the activity. Assessment can be direct or indirect and the chemical species actually detected need not of course be the proteolysis

10 product itself but can for example be a derivative thereof or some further substance.

As used herein, biological activity refers to the *in vivo* activities of a compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological activity, thus, encompasses

15 therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities can be observed in *in vitro* systems designed to test or use such activities. Thus, for purposes herein the biological activity of a luciferase is its oxygenase activity whereby, upon oxidation of a substrate, light is produced.

20 As used herein, functional activity refers to an activity or activities of a polypeptide or portion thereof associated with a full-length (complete) protein. Functional activities include, but are not limited to, biological activity, catalytic or enzymatic activity, antigenicity (ability to bind to or compete with a polypeptide for binding to an anti-polypeptide antibody), immunogenicity, ability to form

25 multimers, and the ability to specifically bind to a receptor or ligand for the polypeptide.

As used herein, a conjugate refers to the compounds provided herein that include one or more SPs, including a CVSP17, particularly single chain protease domains thereof, and one or more targeting agents. These conjugates include

30 those produced by recombinant means as fusion proteins, those produced by chemical means, such as by chemical coupling, through, for example, coupling to sulfhydryl groups, and those produced by any other method whereby at least

-36-

one SP, or a domain thereof, is linked, directly or indirectly via linker(s) to a targeting agent.

As used herein, a targeting agent, is any moiety, such as a protein or effective portion thereof, that provides specific binding of the conjugate to a cell surface receptor, which in some instances can internalize bound conjugates or portions thereof. A targeting agent also can be one that promotes or facilitates, for example, affinity isolation or purification of the conjugate; attachment of the conjugate to a surface; or detection of the conjugate or complexes containing the conjugate.

As used herein, an antibody conjugate refers to a conjugate in which the targeting agent is an antibody.

As used herein, derivative or analog of a molecule refers to a portion derived from or a modified version of the molecule.

As used herein, an effective amount of a compound for treating a particular disease is an amount that is sufficient to ameliorate, or in some manner reduce the symptoms associated with the disease. Such an amount can be administered as a single dosage or can be administered according to a regimen, whereby it is effective. The amount can cure the disease but, typically, is administered in order to ameliorate the symptoms of the disease. Repeated administration can be required to achieve the desired amelioration of symptoms.

As used herein equivalent, when referring to two sequences of nucleic acids, means that the two sequences in question encode the same sequence of amino acids or equivalent proteins. When equivalent is used in referring to two proteins or peptides, it means that the two proteins or peptides have substantially the same amino acid sequence with only amino acid substitutions (such, as but not limited to, conservative changes such as those set forth in Table 1, above) that do not substantially alter the activity or function of the protein or peptide. When equivalent refers to a property, the property does not need to be present to the same extent (*e.g.*, two peptides can exhibit different rates of the same type of enzymatic activity), but the activities are usually substantially the same. Complementary, when referring to two nucleotide sequences, means that the two sequences of nucleotides are capable of

-37-

hybridizing, typically with less than 25%, 15% or 5% mismatches between opposed nucleotides. If necessary, the percentage of complementarity will be specified. Typically the two molecules are selected such that they will hybridize under conditions of high stringency.

5       As used herein, an agent that modulates the activity of a protein or expression of a gene or nucleic acid either decreases or increases or otherwise alters the activity of the protein or, in some manner, up- or down-regulates or otherwise alters expression of the nucleic acid in a cell.

10       As used herein, inhibitor of the activity of an SP encompasses any substances that prohibit or decrease production, post-translational modification(s), maturation, or membrane localization of the SP or any substances that interferes with or decreases the proteolytic efficacy of thereof, particularly of a single chain form in an *in vitro* screening assay.

15       As used herein, a method for treating or preventing neoplastic disease means that any of the symptoms, such as the tumor, metastasis thereof, the vascularization of the tumors or other parameters by which the disease is characterized are reduced, ameliorated, prevented, placed in a state of remission, or maintained in a state of remission. It also means that the hallmarks of neoplastic disease and metastasis can be eliminated, reduced or prevented by  
20       the treatment. Non-limiting examples of the hallmarks include uncontrolled degradation of the basement membrane and proximal extracellular matrix, migration, division, and organization of the endothelial cells into new functioning capillaries, and the persistence of such functioning capillaries.

25       As used herein, pharmaceutically acceptable salts, esters or other derivatives of the conjugates include any salts, esters or derivatives that can be readily prepared by those of skill in this art using known methods for such derivatization and that produce compounds that can be administered to animals or humans without substantial toxic effects and that either are pharmaceutically active or are prodrugs.

30       As used herein, a prodrug is a compound that, upon *in vivo* administration, is metabolized or otherwise converted to the biologically, pharmaceutically or therapeutically active form of the compound. To produce a

-38-

prodrug, the pharmaceutically active compound is modified such that the active compound is regenerated by metabolic processes. The prodrug can be designed to alter the metabolic stability or the transport characteristics of a drug, to mask side effects or toxicity, to improve the flavor of a drug or to alter other

5 characteristics or properties of a drug. By virtue of knowledge of pharmacodynamic processes and drug metabolism *in vivo*, those of skill in this art, once a pharmaceutically active compound is known, can design prodrugs of the compound (see, *e.g.*, Nogrady (1985) *Medicinal Chemistry A Biochemical Approach*, Oxford University Press, New York, pages 388-392).

10 As used herein, a drug identified by the screening methods provided herein refers to any compound that is a candidate for use as a therapeutic or as a lead compound for the design of a therapeutic. Such compounds can be small molecules, including small organic molecules, peptides, peptide mimetics, antisense molecules or dsRNA, such as RNAi, antibodies, fragments of  
15 antibodies, recombinant antibodies and other such compound which can serve as drug candidate or lead compound.

As used herein, a peptidomimetic is a compound that mimics the conformation and certain stereochemical features of the biologically active form of a particular peptide. In general, peptidomimetics are designed to mimic  
20 certain desirable properties of a compound, but not the undesirable properties, such as flexibility, that lead to a loss of a biologically active conformation and bond breakdown. Peptidomimetics may be prepared from biologically active compounds by replacing certain groups or bonds that contribute to the undesirable properties with bioisosteres. Bioisosteres are known to those of  
25 skill in the art. For example the methylene bioisostere  $\text{CH}_2\text{S}$  has been used as an amide replacement in enkephalin analogs (see, *e.g.*, Spatola (1983) pp. 267-357 in Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins, Weinstein, Ed. volume 7, Marcel Dekker, New York). Morphine, which can be administered orally, is a compound that is a peptidomimetic of the peptide  
30 endorphin. For purposes herein, cyclic peptides are included among peptidomimetics.

-39-

As used herein, a promoter region or promoter element refers to a segment of DNA or RNA that controls transcription of the DNA or RNA to which it is operatively linked. The promoter region includes specific sequences that are sufficient for RNA polymerase recognition, binding and transcription initiation.

- 5 This portion of the promoter region is referred to as the promoter. In addition, the promoter region includes sequences that modulate this recognition, binding and transcription initiation activity of RNA polymerase. These sequences can be *cis* acting or can be responsive to *trans* acting factors. Promoters, depending upon the nature of the regulation, can be constitutive or regulated. Exemplary
- 10 promoters contemplated for use in prokaryotes include the bacteriophage T7 and T3 promoters.

- As used herein, a receptor refers to a molecule that has an affinity for a given ligand. Receptors can be naturally-occurring or synthetic molecules. Receptors can also be referred to in the art as anti-ligands. As used herein, the
- 15 receptor and anti-ligand are interchangeable. Receptors can be used in their unaltered state or bound to other polypeptides, including as homodimers. Receptors can be attached to, covalently or noncovalently, or in physical contact with, a binding member, either directly or indirectly via a specific binding substance or linker. Examples of receptors, include, but are not limited to:
- 20 antibodies, cell membrane receptors surface receptors and internalizing receptors, monoclonal antibodies and antisera reactive with specific antigenic determinants [such as on viruses, cells, or other materials], drugs, polynucleotides, nucleic acids, peptides, cofactors, lectins, sugars, polysaccharides, cells, cellular membranes, and organelles.

- 25 Examples of receptors and applications using such receptors, include but are not restricted to:

- a) enzymes: specific transport proteins or enzymes essential to survival of microorganisms, which could serve as targets for antibiotic [ligand] selection;
  - b) antibodies: identification of a ligand-binding site on the antibody
- 30 molecule that combines with the epitope of an antigen of interest can be investigated; determination of a sequence that mimics an antigenic epitope can lead to the development of vaccines of which the immunogen is based on one or

-40-

more of such sequences or lead to the development of related diagnostic agents or compounds useful in therapeutic treatments such as for auto-immune diseases

c) nucleic acids: identification of ligand, such as protein or RNA, binding sites;

5 d) catalytic polypeptides: polymers, including polypeptides, that are capable of promoting a chemical reaction involving the conversion of one or more reactants to one or more products; such polypeptides generally include a binding site specific for at least one reactant or reaction intermediate and an active functionality proximate to the binding site, in which the functionality is  
10 capable of chemically modifying the bound reactant (see, *e.g.*, U.S. Patent No. 5,215,899);

e) hormone receptors: determination of the ligands that bind with high affinity to a receptor is useful in the development of hormone replacement therapies; for example, identification of ligands that bind to such receptors can  
15 lead to the development of drugs to control blood pressure; and

f) opiate receptors: determination of ligands that bind to the opiate receptors in the brain is useful in the development of less-addictive replacements for morphine and related drugs.

As used herein, sample refers to anything which can contain an analyte  
20 for which an analyte assay is desired. The sample can be a biological sample, such as a biological fluid or a biological tissue. Examples of biological fluids include urine, blood, plasma, serum, saliva, semen, stool, sputum, cerebral spinal fluid, tears, mucus, amniotic fluid or the like. Biological tissues are aggregate of cells, usually of a particular kind together with their intercellular substance that  
25 form one of the structural materials of a human, animal, plant, bacterial, fungal or viral structure, including connective, epithelium, muscle and nerve tissues. Examples of biological tissues also include organs, tumors, lymph nodes, arteries and individual cell(s).

As used herein: stringency of hybridization in determining percentage  
30 mismatch is as follows:

1) high stringency: 0.1 x SSPE, 0.1% SDS, 65°C

2) medium stringency: 0.2 x SSPE, 0.1% SDS, 50°C

-41-

3) low stringency: 1.0 x SSPE, 0.1% SDS, 50°C

Those of skill in this art know that the washing step selects for stable hybrids and also know the ingredients of SSPE (see, *e.g.*, Sambrook, E.F. Fritsch, T. Maniatis, in: *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1989), vol. 3, p. B.13, see, also, numerous catalogs that describe commonly used laboratory solutions). SSPE is pH 7.4 phosphate-buffered 0.18 M NaCl. Further, those of skill in the art recognize that the stability of hybrids is determined by  $T_m$ , which is a function of the sodium ion concentration and temperature ( $T_m = 81.5^\circ \text{C} - 16.6(\log_{10}[\text{Na}^+]) + 0.41(\% \text{G} + \text{C}) - 600/l$ ), so that the only parameters in the wash conditions critical to hybrid

It is understood that equivalent stringencies can be achieved using alternative buffers, salts and temperatures. By way of example and not limitation, procedures using conditions of low stringency are as follows (see also Shilo and Weinberg, *Proc. Natl. Acad. Sci. USA* 78:6789-6792 (1981)): Filters containing DNA are pretreated for 6 hours at 40°C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA (10X SSC is 1.5 M sodium chloride, and 0.15 M sodium citrate, adjusted to a pH of 7).

Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 µg/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and  $5-20 \times 10^8$  cpm  $^{32}\text{P}$ -labeled probe is used. Filters are incubated in hybridization mixture for 18-20 hours at 40°C, and then washed for 1.5 hours at 55°C in a solution containing 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 1.5 hours at 60°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and reexposed to film. Other conditions of low stringency which can be used are well known in the art (*e.g.*, as employed for cross-species hybridizations).

By way of example and not way of limitation, procedures using conditions of moderate stringency include, for example, but are not limited to,

-42-

procedures using such conditions of moderate stringency are as follows: Filters containing DNA are pretreated for 6 hours at 55°C in a solution containing 6X SSC, 5X Denhart's solution, 0.5% SDS and 100 µg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution and 5-20 X 10<sup>8</sup> cpm <sup>32</sup>P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 hours at 55°C, and then washed twice for 30 minutes at 60°C in a solution containing 1X SSC and 0.1% SDS. Filters are blotted dry and exposed for autoradiography. Other conditions of moderate stringency which can be used are well-known in the art. Washing of filters is done at 37°C for 1 hour in a solution containing 2X SSC, 0.1% SDS.

By way of example and not way of limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 hours to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 hours at 65°C in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10<sup>8</sup> cpm of <sup>32</sup>P-labeled probe. Washing of filters is done at 37°C for 1 hour in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is followed by a wash in 0.1X SSC at 50°C for 45 minutes before autoradiography. Other conditions of high stringency which can be used are well known in the art.

The term substantially identical or homologous or similar varies with the context as understood by those skilled in the relevant art and generally means at least 60% or 70%, preferably means at least 80%, more preferably at least 90%, and most preferably at least 95% identity.

As used herein, substantially identical to a product means sufficiently similar so that the property of interest is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art

-43-

to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound can, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, target cell refers to a cell that expresses an SP *in vivo*.

As used herein, test substance (or test compound) refers to a chemically defined compound (*e.g.*, organic molecules, inorganic molecules, organic/inorganic molecules, proteins, peptides, nucleic acids, oligonucleotides, lipids, polysaccharides, saccharides, or hybrids among these molecules such as glycoproteins, etc.) or mixtures of compounds (*e.g.*, a library of test compounds, natural extracts or culture supernatants, etc.) whose effect on an SP, particularly a single chain form that includes the protease domain or a sufficient portion thereof for activity, as determined by an *in vitro* method, such as the assays provided herein, is tested.

As used herein, a molecule, such as an antibody, that specifically binds to a polypeptide typically has a binding affinity ( $K_d$ ) of at least about  $10^6$  l/mol,  $10^7$  l/mol,  $10^8$  l/mol,  $10^9$  l/mol,  $10^{10}$  l/mol or greater and binds to a protein of interest generally with at least 2-fold, 5-fold, generally 10-fold or even 100-fold or greater, affinity than to other proteins. For example, an antibody that specifically binds to the protease domain compared to the full-length molecule, such as the zymogen form, binds with at least about 2-fold, typically 5-fold or 10-fold higher affinity, to a polypeptide that contains only the protease domain than to the zymogen form of the full-length. Such specific binding is also referred to as selective binding. Thus, specific or selective binding refers to greater binding affinity (generally at least 1-fold, 2-fold, 5-fold, 10-fold or more) to a targeted site or locus compared to a non-targeted site or locus.

As used herein, the terms a therapeutic agent, therapeutic regimen, radioprotectant, or chemotherapeutic mean conventional drugs and drug

-44-

therapies, including vaccines, which are known to those skilled in the art.

Radiotherapeutic agents are well known in the art.

As used herein, treatment means any manner in which the symptoms of a condition, disorder or disease are ameliorated or otherwise beneficially altered.

5 Treatment also encompasses any pharmaceutical use of the compositions herein.

As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous nucleic acid into cells for either expression or replication thereof. The vectors typically remain episomal, but can be designed to effect integration of a gene or portion thereof into a chromosome of the  
10 genome. Also contemplated are vectors that are artificial chromosomes, such as yeast artificial chromosomes and mammalian artificial chromosomes. Selection and use of such vehicles are well known to those of skill in the art. An expression vector includes vectors capable of expressing DNA that is operatively linked with regulatory sequences, such as promoter regions, that are capable of  
15 effecting expression of such DNA fragments. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those of skill in the art and include those that are  
20 replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

As used herein, protein binding sequence refers to a protein or peptide sequence that is capable of specific binding to other protein or peptide sequences generally, to a set of protein or peptide sequences or to a particular  
25 protein or peptide sequence.

As used herein, epitope tag refers to a short stretch of amino acid residues corresponding to an epitope to facilitate subsequent biochemical and immunological analysis of the epitope tagged protein or peptide. Epitope tagging is achieved by adding the sequence of the epitope tag to a protein-encoding  
30 sequence in an appropriate expression vector. Epitope tagged proteins can be affinity purified using highly specific antibodies raised against the tags.

-45-

As used herein, metal binding sequence refers to a protein or peptide sequence that is capable of specific binding to metal ions generally, to a set of metal ions or to a particular metal ion.

As used herein, a combination refers to any association between two or  
5 among more items.

As used herein, a composition refers to a any mixture. It can be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

As used herein, fluid refers to any composition that can flow. Fluids thus  
10 encompass compositions that are in the form of semi-solids, pastes, solutions, aqueous mixtures, gels, lotions, creams and other such compositions.

As used herein, a cellular extract refers to a preparation or fraction which is made from a lysed or disrupted cell.

As used herein, an agent is said to be randomly selected when the agent  
15 is chosen randomly without considering the specific sequences involved in the association of a protein alone or with its associated substrates, binding partners, etc. An example of randomly selected agents is the use a chemical library or a peptide combinatorial library, or a growth broth of an organism or conditioned medium.

20 As used herein, an agent is said to be rationally selected or designed when the agent is chosen on a non-random basis which takes into account the sequence of the target site and/or its conformation in connection with the agent's action. As described in the Examples, there are proposed binding sites for serine protease and (catalytic) sites in the protein having SEQ ID No. 6.  
25 Agents can be rationally selected or rationally designed by using the peptide sequences that make up these sites.

For clarity of disclosure, and not by way of limitation, the detailed description is divided into the subsections that follow.

-46-

**B. CVSP17 polypeptides, muteins, derivatives and analogs thereof  
SPs**

The serine proteases (SPs) are a family of proteins found in mammals and also other species. SPs share a number of common structural features as  
5 described herein. The proteolytic domains share sequence homology including conserved His, Asp, and Ser residues necessary for catalytic activity that are present in conserved motifs. These SPs are synthesized as zymogens, and activated to two chain forms by specific cleavage.

The SP family can be targeted or therapeutic intervention and also can  
10 serve as diagnostic markers for tumor initiation, development, growth and/or progression. As discussed, members of this family are involved in proteolytic processes that are implicated in tumor development, growth and/or progression. This implication is based upon their functions as proteolytic enzymes in extracellular matrix degradation and remodelling and growth- and pro-angiogenic  
15 factor activation. In addition, their levels of expression or level of activation or their apparent activity resulting from substrate and/or co-factor levels or alterations in substrates and/or co-factors and levels thereof differ in tumor cells from non-tumor cells in the same tissue. Hence, protocols and treatments that alter their activity, such as their proteolytic activities and roles in signal  
20 transduction, and/or their expression, such as by contacting them with a compound that modulates their activity and/or expression, could impact tumor development, growth and/or progression. Also, in some instances, the level of activation and/or expression can be altered in tumors, such as pancreas, stomach, uterus, lung, colon and cervical cancers, and also breast, prostate or  
25 leukemias. The SP, thus, can serve as a diagnostic marker for tumors.

In other instances the SP protein can exhibit altered activity by virtue of a change in activity or expression of a co-factor therefor or a substrate therefor. Detection of the SPs, particularly the protease domains, in body fluids, such as serum, blood, saliva, cerebral spinal fluid, synovial fluid and interstitial fluids,  
30 urine, sweat and other such fluids and secretions, can serve as a diagnostic tumor marker. In particular, detection of higher levels of such polypeptides in a subject compared to a subject known not to have any neoplastic disease or

-47-

compared to earlier samples from the same subject, can be indicative of neoplastic disease in the subject.

Provided is a family member designated CVSP17. The CVSP17s provided herein are serine proteases that are expressed and/or activated in certain tumors; hence their activation or expression can serve as a diagnostic marker for tumor development, growth and/or progression. The CVSP17 is also provided for use as a drug target and used in screening assays, including those exemplified herein. The single chain proteolytic domain can function *in vitro* and, hence is useful in *in vitro* assays for identifying agents that modulate the activity of members of this family. In addition the two-chain form or the an activated full-length or truncated forms thereof, such as forms in which the signal peptide is removed, can also be used in such assays. Assays for activation also are provided.

In certain embodiments, the CVSP17 polypeptide is detectable in a body fluid at a level that differs from its level in body fluids in a subject not having a tumor. In other embodiments, the polypeptide is present in a tumor; and a substrate or cofactor for the polypeptide is expressed at levels that differ from its level of expression in a non-tumor cell in the same type of tissue.

#### CVSP17

Provided are substantially purified CVSP17 zymogens, activated two chain forms, single chain protease domains and two chain protease domains. A full-length CVSP17 polypeptide, including the signal sequence, is set forth in SEQ ID No. 6 The signal sequence can be cleaved upon expression or the encoding nucleic acid can be deleted prior to expression.

Also provided is a substantially purified protein including a sequence of amino acids that has at least 60%, 70%, 80%, 90% or about 95%, identity to the CVSP17 where the percentage identity is determined using standard algorithms and gap penalties that maximize the percentage identity. A human CVSP17 polypeptide is exemplified, although other mammalia CVSP17 polypeptides are contemplated. Splice variants of the CVSP17, particularly those with a proteolytically active protease domain, are contemplated herein.

-48-

In other embodiments, substantially purified polypeptides that include a protease domain of a CVSP17 polypeptide or a catalytically active portion thereof, but that do not include the entire sequence of amino acids set forth in SEQ ID No. 6 are provided. Among these are polypeptides that include a  
5 sequence of amino acids that has at least 60%, 70%, 80%, 90%, 95% or 100% sequence identity to SEQ ID No. 6.

Provided are substantially purified CVSP17 polypeptides and functional domains thereof, including catalytically active domains and portions, that have at least about 60%, 70%, 80%, 90% or about 95% sequence identity with a  
10 protease domain that includes the sequence of amino acids set forth in SEQ ID No. 6 or a catalytically active portion thereof.

With reference to SEQ ID No. 6, the protease activation cleavage site is between R<sub>104</sub> and I<sub>105</sub>; the catalytic triad H<sub>145</sub>, D<sub>191</sub> and S<sub>286</sub> occur in 3 highly-conserved regions of the catalytic domain. There is a potential N-glycosylation  
15 site (...N<sub>97</sub>VT...). The following cysteine pairings in the protease domain are noted: C<sub>130</sub>-C<sub>146</sub>, C<sub>225</sub>-C<sub>292</sub>, C<sub>256</sub>-C<sub>271</sub> and C<sub>282</sub>-C<sub>313</sub>. Cys pairing is predicted to be between C<sub>88</sub>-C<sub>211</sub>, which links the protease domain to the remainder of the polypeptide). Hence C<sub>211</sub> is a free Cys in the protease domain, which also can be provided as a two chain molecule. The single chain forms of the protease  
20 domain is proteolytically active.

Also provided are polypeptides that are encoded by the nucleic acid molecules provided herein. Included among those polypeptides are the CVSP17 protease domain or a polypeptide with amino acid changes such that the specificity and protease activity is not eliminated and is retained at least 1%,  
25 2%, 3%, 5%, 10%, 20%, 30%, 40%, 50% or remains substantially unchanged or increases. In particular, a substantially purified mammalian SP protein is provided that includes a serine protease catalytic domain and can additionally include other domains. The CVSP17 can form homodimers and can also form heterodimers with some other protein, such as a membrane-bound protein.

30 Domains, fragments, derivatives or analogs of a CVSP17 that are functionally active are capable of exhibiting one or more functional activities

-49-

associated with a CVSP17 polypeptide, such as serine protease activity, immunogenicity and antigenicity, are provided.

Antigenic epitopes that contain at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 40, 50, and typically 10-15 amino acids of the CVSP17 polypeptide, particularly those from the C-terminus (after the protease domain) are provided. These antigenic epitopes are used, for example, to raise antibodies. Antibodies specific for each epitope or combinations thereof and for single and/or two-chain forms are also provided. Antibodies that specifically bind to the active site region of a zymogen and activated form are provided.

#### 10                   **Muteins and derivatives of CVSP17 polypeptides**

Full-length CVSP17, zymogen and activated forms thereof and CVSP17 protease domains, portions thereof, and muteins and derivatives of such polypeptides are provided. Among the derivatives are those based on animal CVSP17s, including, but are not limited to, rodent, such as mouse and rat; fowl, 15 such as chicken; ruminants, such as goats, cows, deer, sheep; ovine, such as pigs; and humans. For example, CVSP17 derivatives can be made by altering their sequences by substitutions, additions or deletions. CVSP17 derivatives include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of CVSP17, including altered 20 sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence can be 25 selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include 30 arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid (see, *e.g.*, Table 1). Muteins of the CVSP17 or a domain thereof, such as a protease domain, in which up to about 10%, 20%,

-50-

30%, 40%, 50%, 60%, 70%, 80%, 85%, 90% or 95% of the amino acids are replaced with another amino acid are provided. Generally such muteins retain at least about 1%, 2%, 3%, 5%, 7%, 8%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or more (or in increased activity, *i.e.*, 101, 102, 103, 5 104, 105, 110% or greater) of the protease activity of the unmutated protein. Those of skill in the art recognize that a polypeptide that retains at least 1% of the activity of the wild-type protease is sufficiently active for use in screening assays or in other applications.

Muteins in which one or more of the Cys residues, particularly, a residue 10 that is paired in the activated two-chain form, but unpaired in the protease domain alone (*i.e.*, the Cys at residue position 211 (see SEQ ID Nos. 5 and 6) in the protease domain), is/are replaced with any amino acid, typically, although not necessarily, a conservative amino acid residue, such as Ser, are contemplated. Muteins in which 10%, 20%, 30%, 35%, 40%, 45%, 50% or more of the amino 15 acids are replaced but the resulting polypeptide retains at least about 1%, 3, 5%, 10%, 20%, 30%, 35%, 40%, 45%, 50%, 60%, 70%, 80%, 90% or 95% of the catalytic activity as the unmodified form for the same substrate also are provided.

#### Protease domains

Isolated, substantially pure polypeptides that include the protease domains 20 or catalytically active portions thereof as single chain forms of SPs are provided. The protease domains can be included in a longer protein, and such longer proteins include the full-length CVSP17 zymogen. Provided herein are isolated substantially pure polypeptides that contain the protease domain of a CVSP17 as a single chain. The CVSP17 provided herein is expressed or activated by or in 25 tumor cells, typically at a level that differs from the level in which they are expressed by the non-tumor cell of the same type. Hence, for example, if the SP expressed by a prostate or ovarian tumor cell is to be of interest herein with respect to ovarian or prostate cancer, it should have an expression, extent of activation or activity that is different from that in non-tumor cells. CVSP17 is 30 expressed in lung, colon, prostate, breast, uterine, ovarian and other tumor cells.

SP protease domains include the single chain protease domains of CVSP17. Provided are the protease domains or proteins that include a portion of

-51-

an SP that is the protease domain of any SP, particularly a CVSP17. The protein can also include other non-SP sequences of amino acids, but includes the protease domain or a sufficient portion thereof to exhibit catalytic and/or binding activity in any *in vitro* assay that assesses such activity(ies), such as any provided  
5 herein. Also provided are two chain activated forms of the full length protease and also two chain forms of the protease domain.

In an embodiment, the substantially purified SP protease is encoded by a nucleic acid that hybridizes to a nucleic acid molecule containing the protease domain encoded by the nucleotide sequence set forth in SEQ. ID No. 5 under at  
10 least moderate, generally high, stringency conditions, such that the protease domain encoding nucleic acid thereof hybridizes along its full length or along at least about 70%, 80% or 90% of the full length. In other embodiments the substantially purified SP protease is a single chain polypeptide that includes substantially the sequence of amino acids set forth as amino acids 105-332 in  
15 SEQ ID No. 6, or a catalytically active portion thereof. Polypeptides that additionally include amino acids at the C-terminus, such as all or a portion of the 303-amino acids following the protease domain (aa 333 aa 635 in SEQ ID No. 6) in the exemplified embodiment are provided. Dimers and other multimers of the full length and catalytically active portions of the polypeptides  
20 that include at least amino acids 333-427, such as 333-453 (or equivalent regions in other embodiments) are provided.

A signal peptide (amino acids 1-17 of SEQ ID No. 6 in the exemplified embodiment) is also provided. In addition the mature CVSP17 polypeptide with the signal sequence removed and catalytically active portions thereof, including  
25 those that include all or a portion of the C-terminus beyond the protease domain are provided.

As described below, all forms of the CVSP17, including the pro-polypeptide with the signal sequence, the mature polypeptide and catalytically active portions thereof, the protease domains and catalytically active portions  
30 thereof, two-chain and single chain forms of any of these proteins are provided herein and can be used in the screening assays and for preparing specific

-52-

antibodies therefor. The expression, quantity and/or activation of the protein in tumor cells and body fluids can be diagnostic of disease or its absence.

**Nucleic acid molecules, vectors and plasmids, cells and expression of CVSP17 polypeptides**

**5 Nucleic acid molecules**

Due to the degeneracy of nucleotide coding sequences, other nucleic sequences which encode substantially the same amino acid sequence as a CVSP17 gene can be used. These include but are not limited to nucleotide sequences comprising all or portions of CVSP17 genes that are altered by the  
10 substitution of different codons that encode the amino acid residue within the sequence, thus producing a silent change.

Also provided are nucleic acid molecules that hybridize to the above-noted sequences of nucleotides encoding CVSP17 at least at low stringency, at moderate stringency, and/or at high stringency, and that encode the protease  
15 domain and/or the full length protein or other domains of a CVSP17 or a splice variant or allelic variant thereof. Generally the molecules hybridize under such conditions along their full length (or along at least about 70%, 80% or 90% of the full length) for at least one domain and encode at least one domain, such as the protease domain, of the polypeptide. In particular, such nucleic acid  
20 molecules include any isolated nucleic fragment that encodes at least one domain of a serine protease, that (1) contains a sequence of nucleotides that encodes the protease or a functionally active, such as catalytically active, domain thereof, and (2) is selected from among sequences of nucleic acids that encode a CVSP17 polypeptide that is:

25 a polypeptide encoded by the sequence of nucleotides set forth in SEQ ID No. 5;

a polypeptide encoded by a sequence of nucleotides that hybridizes under conditions of low, moderate or high stringency to the sequence of nucleotides set forth in SEQ ID No. 5;

30 a polypeptide that comprises the sequence of amino acids set forth in SEQ ID No. 6;

-53-

a polypeptide that comprises a sequence of amino acids having at least about 60%, 70%, 80%, 90% or about 95% sequence identity with the sequence of amino acids set forth in SEQ ID No. 6; and/or

a polypeptide encoded by a splice variant of a sequence of  
5 nucleotides that encodes a CVSP17.

In particular, the nucleic acid molecules include CVSP17 polypeptides, as that include a protease domain of serine protease 17 (CVSP17) or a catalytically active portion thereof, where:

a) the CVSP17 polypeptide also includes at least 10 or more  
10 contiguous amino acids from residues 397-427 of SEQ ID No. 6 or comprises 10 or more contiguous amino acids encoded by a sequence of nucleotides that hybridizes under conditions of high stringency to a sequence of nucleotides that encodes residues 397-427 of SEQ ID No. 6; or

b) the CVSP17 portion of the polypeptide is only the protease domain  
15 of a CVSP17 or a catalytically active portion thereof, except that the protease domain does not include Cys Arg Ser Thr Arg Ser (SEQ ID No. 18) as a contiguous sequence;

c) the CVSP17 polypeptide contains only residues 19-332 of SEQ ID  
No. 6;

20 d) the CVSP17 polypeptide contains the sequence of amino acids set forth in SEQ ID No. 6;

e) the CVSP17 polypeptide is encoded by a sequence of nucleotides that hybridizes under conditions of at least moderate, and can be high, stringency along at least 70% of its full length to a sequence of nucleotides than  
25 encodes a polypeptide of any of a)-e); and/or

f) the CVSP17 polypeptide has at least 60%, 60%, 70%, 80%, 90% or about 95% sequence identity with the sequence identity with a polypeptide of any of a)-e). Smaller portions thereof that retain protease activity are contemplated.

-54-

Among these nucleic acid molecules are those that contain:

- (a) a sequence of nucleotides that includes a sequence of nucleotides set forth in SEQ ID Nos. 5
- (b) a sequence of nucleotides that encodes such portion or the full length CVSP17 protease, as defined herein, and hybridizes under conditions of moderate or high stringency to nucleic acid that is complementary to an mRNA transcript present in a mammalian cell that encodes a CVSP17 polypeptide or catalytically active fragment thereof;
- (c) a sequence of nucleotides that encodes a CVSP17 protease, as defined herein, or a catalytically active portion thereof that includes a sequence of amino acids encoded by such portion or a full length open reading frame;
- (d) a sequence of nucleotides that encodes the serine protease that includes a sequence of amino acids encoded by a sequence of nucleotides that encodes the protease and hybridizes under conditions of high stringency to DNA that is complementary to the mRNA transcript of (b);
- (e) a sequence of nucleotides that encodes a splice variant of any of (a)-(d);
- (f) a sequence of nucleotides that encodes a CVSP17 that has at least 1% of the catalytic activity of the CVSP17 of SEQ ID No. 6 and has at least 60%, 70%, 80%, 90% or 95% sequence identity with the CVSP17 of SEQ ID No. 6. and
- (g) a sequence of nucleotides that includes degenerate codons of all or a portion of any of (a)-(f).

The isolated nucleic acid fragment is DNA, including genomic or cDNA, or is RNA, or can include other components, such as peptide nucleic acid (PNA).

The isolated nucleic acid can include additional components, such as

- heterologous or native promoters, and other transcriptional and translational regulatory sequences, these genes can be linked to other genes, such as reporter genes or other indicator genes or genes that encode indicators.

-55-

The CVS17s provided herein are encoded by a nucleic acid that includes sequence encoding a protease domain that exhibits proteolytic and/or binding activity and that hybridizes to a nucleic acid molecule including the sequence of nucleotides set forth in SEQ ID No. 5, typically under moderate, generally under  
5 high stringency, conditions and generally along the full length of the protease domain or along at least about 70%, 80% or 90% of the full length. Splice variants are also provided herein.

In a specific embodiment, a nucleic acid that encodes a CVSP, designated CVSP17 is provided. In particular, the nucleic acid includes the sequence of  
10 nucleotides set forth in SEQ ID No. 5 or a portion thereof that encodes a catalytically active polypeptide. Also provided are nucleic acid molecules that hybridize under conditions of at least low stringency, generally moderate stringency, more typically high stringency to the SEQ ID No. 5 or degenerates thereof.

15 In one embodiment, the isolated nucleic acid fragment hybridizes to a nucleic acid molecule containing the nucleotide sequence set forth in SEQ ID No: 5 (or degenerates thereof) under high stringency conditions, while another embodiment contains the sequence of nucleotides set forth in SEQ ID Nos. 5. A full-length CVSP17 is set forth in SEQ ID No. 6 and is encoded by SEQ ID No. 5  
20 or degenerates thereof.

Also contemplated are nucleic acid molecules that encode a single chain SP protease that have proteolytic activity in an *in vitro* proteolysis assay and that have at least 60%, 70%, 80%, 85%, 90% or 95% sequence identity with the full length of a protease domain of a CVSP17 polypeptide, or that hybridize  
25 along their full length or along at least about 70%, 80% or 90% of the full length to a nucleic acid that encodes a protease domain, particularly under conditions of moderate, generally high, stringency.

The isolated nucleic acids can contain least 10 nucleotides, 25 nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, or 200  
30 nucleotides or more contiguous nucleotides of a CVSP17-encoding sequence, or a full-length SP coding sequence. In another embodiment, the nucleic acids are smaller than 35, 200 or 500 nucleotides in length. Nucleic acids that hybridize

-56-

to or are complementary to a CVSP17-encoding nucleic acid molecule can be single or double-stranded. For example, nucleic acids are provided that include a sequence complementary to (specifically are the inverse complement of) at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of a CVSP17  
5 encoding nucleic acid, particularly the protease domain thereof. For CVSP17 the full-length protein or a domain or active fragment thereof is also provided.

For each of the nucleic acid molecules, the nucleic acid can be DNA or RNA or PNA or other nucleic acid analogs or can include non-natural nucleotide bases. Also provided are isolated nucleic acid molecules that include a sequence  
10 of nucleotides complementary to the nucleotide sequence encoding an SP.

**Probes, primers, antisense oligonucleotides and dsRNA**

Also provided are fragments thereof or oligonucleotides that can be used as probes or primers and that contain at least about 10, 14, 16 nucleotides, generally less than 1000 or less than or equal to 100, set forth in SEQ ID No. 5  
15 (or the complement thereof); or contain at least about 30 nucleotides (or the complement thereof) or contain oligonucleotides that hybridize along their full length or along at least about 70%, 80% or 90% of the full length to any such fragments or oligonucleotides. The length of the fragments is a function of the purpose for which they are used and/or the complexity of the genome of  
20 interest. Generally probes and primers contain less than about 500, 150, 100 nucleotides.

Probes and primers derived from the nucleic acid molecules are provided. Such probes and primers contain at least 8, 14, 16, 30, 100 or more contiguous nucleotides with identity to contiguous nucleotides of a CVSP17. The probes  
25 and primers are optionally labeled with a detectable label, such as a radiolabel or a fluorescent tag, or can be mass differentiated for detection by mass spectrometry or other means.

Also provided is an isolated nucleic acid molecule that includes the sequence of molecules that is complementary to the nucleotide sequence  
30 encoding CVSP17 or the portion thereof. Double-stranded RNA (dsRNA), such as RNAi is also provided.

**Plasmids, vectors and cells**

-57-

Plasmids and vectors containing the nucleic acid molecules are also provided. Cells containing the vectors, including cells that express the encoded proteins are provided. The cell can be a bacterial cell, a yeast cell, a fungal cell, a plant cell, an insect cell or an animal cell. Methods for producing an SP or  
5 single chain form of the protease domain thereof by, for example, growing the cell under conditions whereby the encoded SP is expressed by the cell, and recovering the expressed protein, are provided herein. As noted, for CVSP17, the full-length zymogens and activated proteins and activated (two chain) protease and single chain protease domains are provided.

10 As discussed below, the CVSP17 polypeptide, and catalytically active portions thereof, can be expressed as a secreted protein using the native signal sequence or a heterologous signal. Alternatively the protein can be expressed as inclusion bodies in the cytoplasm and isolated therefrom. The resulting protein can be treated to refold (see, *e.g.*, EXAMPLE 1). Active protease domain can be  
15 produced by expression in inclusion bodies, isolation therefrom and denaturation followed by refolding.

**C. Tumor specificity and tissue expression profiles**

Each SP has a characteristic tissue expression profile; the SPs in particular, although not exclusively expressed or activated in tumors, exhibit  
20 characteristic tumor tissue expression or activation profiles. In some instances, SPs can have different activity in a tumor cell from a non-tumor cell by virtue of a change in a substrate or cofactor therefor or other factor that would alter the apparent functional activity of the SP. Hence each can serve as a diagnostic marker for particular tumors, by virtue of a level of activity and/or expression or  
25 function in a subject (i.e. a mammal, particularly a human) with neoplastic disease, compared to a subject or subjects that do not have the neoplastic disease. In addition, detection of activity (and/or expression) in a particular tissue can be indicative of neoplastic disease.

Circulating SPs in body fluids can be indicative of neoplastic disease.  
30 Secreted CVSP17 or activated CVSP17 is indicative of neoplastic disease. Also, by virtue of the activity and/or expression profiles of each SP, they can serve as therapeutic targets, such as by administration of modulators of the activity

-58-

thereof, or, as by administration of a prodrug specifically activated by one of the SPs.

#### **Tissue expression profiles**

##### **CVSP17**

5 CVSP17 is expressed in cervical, colon and pancreatic tumors.  
In particular, CVSP17 is strongly expressed in cervical carcinoma cells. It or a fragment or splice variant thereof is expressed or activated in colon tumors and also in pancreatic tumors.

10 Its expression or activation in certain cells, such as cervical cells, can serve as a tumor marker; whereas in other tissues, the absence of expression or activation, can serve as a tumor marker. Detection of CVSP17 in a body fluid also can be indicative of a tumor.

#### **D. Identification and isolation of SP protein genes**

The SP polypeptides, including CVSP17 polypeptides, or domains thereof,  
15 can be obtained by methods well known in the art for protein purification and recombinant protein expression. Any method known to those of skill in the art for identification of nucleic acids that encode desired genes can be used. Any method available in the art can be used to obtain a full length (*i.e.*, encompassing the entire coding region) cDNA or genomic DNA clone encoding  
20 an SP protein. In particular, the polymerase chain reaction (PCR) can be used to amplify a sequence identified as being differentially expressed or encoding proteins activated at different levels in tumor and non-tumor cells or tissues, *e.g.*, nucleic acids encoding a CVSP17 polypeptide (SEQ. NOs: 5, 6, 12 and 13), in a genomic or cDNA library. Oligonucleotide primers that hybridize to  
25 sequences at the 3' and 5' termini of the identified sequences can be used as primers to amplify by PCR sequences from a nucleic acid sample (RNA or DNA), typically a cDNA library, from an appropriate source (*e.g.*, tumor or cancer tissue).

30 Amplification, such PCR can be carried out by a thermal cycler and Taq polymerase (Gene Amp<sup>™</sup>). The amplified nucleic acid can include mRNA or cDNA or genomic DNA from any eukaryotic species. One can choose to synthesize several different degenerate primers, for use in the PCR reactions. It is also

-59-

possible to vary the stringency of hybridization conditions used in priming the PCR reactions, to amplify nucleic acid orthologs or homologs (*e.g.*, to obtain *SP protein* sequences from species other than humans or to obtain human sequences with homology to CVSP17 polypeptide) by allowing for greater or

5 lesser degrees of nucleotide sequence similarity between the known nucleotide sequence and the nucleic acid homolog being isolated. For cross species hybridization, low or moderate stringency conditions are used. For same species hybridization, moderate or high stringency conditions generally are used. After successful amplification of the nucleic acid containing all or a portion of the

10 identified SP protein sequence or of a nucleic acid encoding all or a portion of an SP protein homolog, that segment can be molecularly cloned and sequenced, and used as a probe to isolate a complete cDNA or genomic clone. This, in turn, permits the determination of the gene's complete nucleotide sequence, the analysis of its expression, and the production of its protein product for functional

15 analysis. Once the nucleotide sequence is determined, an open reading frame encoding the SP protein gene protein product can be determined by any method well known in the art for determining open reading frames, for example, using publicly available computer programs for nucleotide sequence analysis. Once an open reading frame is defined, it is routine to determine the amino acid sequence

20 of the protein encoded by the open reading frame. In this way, the nucleotide sequences of the entire SP protein genes as well as the amino acid sequences of SP proteins and analogs can be identified.

Any eukaryotic cell potentially can serve as the nucleic acid source for the molecular cloning of the SP protein gene. The nucleic acids can be isolated

25 from vertebrate, mammalian, human, porcine, bovine, feline, avian, equine, canine, as well as additional primate sources, insects, plants, etc. The DNA can be obtained by standard procedures known in the art from cloned DNA (*e.g.*, a DNA "library"), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from the desired cell (see, for

30 example, Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, *DNA Cloning: A Practical Approach*, MRL Press, Ltd., Oxford,

-60-

U.K. Vol. I, II). Clones derived from genomic DNA can contain regulatory and intron DNA regions in addition to coding regions; clones derived from cDNA contains only exon sequences. Whatever the source, the gene should be molecularly cloned into a suitable vector for propagation of the gene.

5 In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which encode the desired gene. The DNA can be cleaved at specific sites using various restriction enzymes. Alternatively, one can use DNase in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, for example, by sonication. The linear DNA  
10 fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene can be accomplished in a number of ways.  
15 For example, a portion of the SP protein (of any species) gene (*e.g.*, a PCR amplification product obtained as described above or an oligonucleotide having a sequence of a portion of the known nucleotide sequence) or its specific RNA, or a fragment thereof be purified and labeled, and the generated DNA fragments can be screened by nucleic acid hybridization to the labeled probe (Benton and  
20 Davis, *Science* 196:180 (1977); Grunstein and Hogness, *Proc. Natl. Acad. Sci. U.S.A.* 72:3961 (1975)). Those DNA fragments with substantial homology to the probe hybridize. It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of fragment sizes with those expected according to a known restriction map if such is available or by DNA  
25 sequence analysis and comparison to the known nucleotide sequence of SP protein. Further selection can be carried out on the basis of the properties of the gene. Alternatively, the presence of the gene can be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper  
30 mRNA, can be selected which produce a protein that, *e.g.*, has similar or identical electrophoretic migration, isoelectric focusing behavior, proteolytic digestion maps, antigenic properties, serine protease activity. If an anti-SP

-61-

protein antibody is available, the protein can be identified by binding of labeled antibody to the putatively SP protein synthesizing clones, in an ELISA (enzyme-linked immunosorbent assay)-type procedure.

Alternatives to isolating the CVSP17 polypeptide genomic DNA include, but are not limited to, chemically synthesizing the gene sequence from a known sequence or making cDNA to the mRNA that encodes the SP protein. For example, RNA for cDNA cloning of the SP protein gene can be isolated from cells expressing the protein. The identified and isolated nucleic acids can then be inserted into an appropriate cloning vector. A large number of vector-host systems known in the art can be used. Possible vectors include, but are not limited to, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Such vectors include, but are not limited to, bacteriophages such as lambda derivatives, or plasmids such as pBR322 or pUC plasmid derivatives or the Bluescript vector (Stratagene, La Jolla, CA). The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. If the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules can be enzymatically modified. Alternatively, any site desired can be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers can comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and SP protein gene can be modified by homopolymeric tailing. Recombinant molecules can be introduced into host cells via, for example, transformation, transfection, infection, electroporation and sonoporation, so that many copies of the gene sequence are generated.

In specific embodiments, transformation of host cells with recombinant DNA molecules that incorporate the isolated SP protein gene, cDNA, or synthesized DNA sequence enables generation of multiple copies of the gene. Thus, the gene can be obtained in large quantities by growing transformants, isolating the recombinant DNA molecules from the transformants and, when necessary, retrieving the inserted gene from the isolated recombinant DNA.

-62-

**E. Vectors, plasmids and cells that contain nucleic acids encoding an SP protein or protease domain thereof and expression of SP proteins**

**Vectors and cells**

For recombinant expression of one or more of the SP proteins, the nucleic acid containing all or a portion of the nucleotide sequence encoding the SP protein can be inserted into an appropriate expression vector, *i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted protein coding sequence. The necessary transcriptional and translational signals also can be supplied by the native promoter for SP genes, and/or their flanking regions.

Also provided are vectors that contain nucleic acid encoding the SPs. Cells containing the vectors are also provided. The cells include eukaryotic and prokaryotic cells, and the vectors are any suitable for use therein.

Prokaryotic and eukaryotic cells, including endothelial cells, containing the vectors are provided. Such cells include bacterial cells, yeast cells, fungal cells, Archea, plant cells, insect cells and animal cells. The cells are used to produce an SP protein or protease domain thereof by growing the above-described cells under conditions whereby the encoded SP protein or protease domain of the SP protein is expressed by the cell, and recovering the expressed protease domain protein. For purposes herein, the protease domain can be secreted into the medium.

In one embodiment, the vectors include a sequence of nucleotides that encodes a polypeptide that has protease activity and contains all or a portion of the protease domain, or multiple copies thereof, of an SP protein are provided. Also provided are vectors that comprise a sequence of nucleotides that encodes the protease domain and additional portions of an SP protein up to and including a full length SP protein, as well as multiple copies thereof, are also provided. The vectors can selected for expression of the SP protein or protease domain thereof in the cell or such that the SP protein is expressed as a secreted protein. When the protease domain is expressed the nucleic acid is linked to nucleic acid encoding a secretion signal, such as the *Saccharomyces cerevisiae*  $\alpha$  mating factor signal sequence or a portion thereof, or the native signal sequence.

-63-

A variety of host-vector systems can be used to express the protein coding sequence. These include but are not limited to mammalian cell systems infected with virus (*e.g.* vaccinia virus, adenovirus, etc.); insect cell systems infected with virus (*e.g.* baculovirus); microorganisms such as yeast containing  
5 yeast vectors; or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their strengths and specificities. Depending on the host-vector system used, any one of a number of suitable transcription and translation elements can be used.

Any methods known to those of skill in the art for the insertion of DNA  
10 fragments into a vector can be used to construct expression vectors containing a chimeric gene containing of appropriate transcriptional/translational control signals and protein coding sequences. These methods can include *in vitro* recombinant DNA and synthetic techniques and *in vivo* recombinants (genetic recombination). Expression of nucleic acid sequences encoding SP protein, or  
15 domains, derivatives, fragments or homologs thereof, can be regulated by a second nucleic acid sequence so that the genes or fragments thereof are expressed in a host transformed with the recombinant DNA molecule(s). For example, expression of the proteins can be controlled by any promoter/enhancer known in the art. In a specific embodiment, the promoter is not native to the  
20 genes for SP protein. Promoters which can be used include but are not limited to the SV40 early promoter (Bernoist and Chambon, *Nature* 290:304-310 (1981)), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto *et al.* *Cell* 22:787-797 (1980)), the herpes thymidine kinase promoter (Wagner *et al.*, *Proc. Natl. Acad. Sci. USA* 78:1441-1445 (1981)), the  
25 regulatory sequences of the metallothionein gene (Brinster *et al.*, *Nature* 296:39-42 (1982)); prokaryotic expression vectors such as the  $\beta$ -lactamase promoter (Villa-Kamaroff *et al.*, *Proc. Natl. Acad. Sci. USA* 75:3727-3731 1978)) or the *tac* promoter (DeBoer *et al.*, *Proc. Natl. Acad. Sci. USA* 80:21-25 (1983)); see also "Useful Proteins from Recombinant Bacteria": in Scientific American  
30 242:79-94 (1980)); plant expression vectors containing the nopaline synthetase promoter (Herrar-Estrella *et al.*, *Nature* 303:209-213 (1984)) or the cauliflower mosaic virus 35S RNA promoter (Gardner *et al.*, *Nucleic Acids Res.* 9:2871

-64-

(1981)), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., *Nature* 310:115-120 (1984)); promoter elements from yeast and other fungi such as the Gal4 promoter, the alcohol dehydrogenase promoter, the phosphoglycerol kinase promoter, the alkaline phosphatase promoter, and the following animal transcriptional control regions that exhibit tissue specificity and have been used in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., *Cell* 38:639-646 (1984); Ornitz et al., *Cold Spring Harbor Symp. Quant. Biol.* 50:399-409 (1986); MacDonald, *Hepatology* 7:425-515 (1987)); insulin gene control region which is active in pancreatic beta cells (Hanahan et al., *Nature* 315:115-122 (1985)), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., *Cell* 38:647-658 (1984); Adams et al., *Nature* 318:533-538 (1985); Alexander et al., *Mol. Cell Biol.* 7:1436-1444 (1987)), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., *Cell* 45:485-495 (1986)), albumin gene control region which is active in liver (Pinckert et al., *Genes and Devel.* 1:268-276 (1987)), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., *Mol. Cell Biol.* 5:1639-1648 (1985); Hammer et al., *Science* 235:53-58 (1987)), alpha-1 antitrypsin gene control region which is active in liver (Kelsey et al., *Genes and Devel.* 1:161-171 (1987)), beta globin gene control region which is active in myeloid cells (Mogam et al., *Nature* 315:338-340 (1985); Kollias et al., *Cell* 46:89-94 (1986)), myelin basic protein gene control region which is active in oligodendrocyte cells of the brain (Readhead et al., *Cell* 48:703-712 (1987)), myosin light chain-2 gene control region which is active in skeletal muscle (Sani, *Nature* 314:283-286 (1985)), and gonadotrophic releasing hormone gene control region which is active in gonadotrophs of the hypothalamus (Mason et al., *Science* 234:1372-1378 (1986)).

In a specific embodiment, a vector is used that contains a promoter operably linked to nucleic acids encoding an SP protein, or a domain, fragment, derivative or homolog, thereof, one or more origins of replication, and optionally, one or more selectable markers (e.g., an antibiotic resistance gene). Expression vectors containing the coding sequences, or portions thereof, of an SP protein,

-65-

are made, for example, by subcloning the coding portions into the EcoRI restriction site of each of the three pGEX vectors (glutathione S-transferase expression vectors (Smith and Johnson, *Gene* 7:31-40 (1988)). This allows for the expression of products in the correct reading frame. Vectors and systems  
5 for expression of the protease domains of the SP proteins include the well known *Pichia* vectors (available, for example, from Invitrogen, San Diego, CA), particularly those designed for secretion of the encoded proteins. One exemplary vector is described in the EXAMPLES.

Plasmids for transformation of *E. coli* cells, include, for example, the pET  
10 expression vectors (see, U.S patent 4,952,496; available from NOVAGEN, Madison, WI; see, also literature published by Novagen describing the system). Such plasmids include pET 11a, which contains the T7lac promoter, T7 terminator, the inducible *E. coli* lac operator, and the lac repressor gene; pET 12a-c, which contains the T7 promoter, T7 terminator, and the *E. coli* ompT  
15 secretion signal; and pET 15b and pET19b (NOVAGEN, Madison, WI), which contain a His-Tag™ leader sequence for use in purification with a His column and a thrombin cleavage site that permits cleavage following purification over the column; the T7-lac promoter region and the T7 terminator.

The vectors are introduced into host cells, such as *Pichia* cells and  
20 bacterial cells, such as *E. coli*, and the proteins expressed therein. *Pichia* strains, which are known and readily available, include, for example, GS115. Bacterial hosts can contain chromosomal copies of DNA encoding T7 RNA polymerase operably linked to an inducible promoter, such as the lacUV promoter (see, U.S. Patent No. 4,952,496). Such hosts include, but are not  
25 limited to, the lysogenic *E. coli* strain BL21(DE3).

#### Expression and production of proteins

The SP domains, derivatives and analogs can be produced by various methods known in the art. For example, once a recombinant cell expressing an SP protein, or a domain, fragment or derivative thereof, is identified, the  
30 individual gene product can be isolated and analyzed. This is achieved by assays based on the physical and/or functional properties of the protein, including, but not limited to, radioactive labeling of the product followed by

-66-

analysis by gel electrophoresis, immunoassay, cross-linking to marker-labeled product.

The CVSP17 polypeptides can be isolated and purified by standard methods known in the art (either from natural sources or recombinant host cells  
5 expressing the complexes or proteins), including but not restricted to column chromatography (*e.g.*, ion exchange, affinity, gel exclusion, reversed-phase high pressure and fast protein liquid), differential centrifugation, differential solubility, or by any other standard technique used for the purification of proteins. Functional properties can be evaluated using any suitable assay known in the  
10 art.

Alternatively, once an SP protein or its domain or derivative is identified, the amino acid sequence of the protein can be deduced from the nucleotide sequence of the gene which encodes it. In addition, domains, analogs and derivatives of an SP protein can be chemically synthesized by standard chemical  
15 methods known in the art (*e.g.* see Hunkapiller *et al.* (1984) *Nature* 310:105-111). For example, a peptide corresponding to a portion of an SP protein, which includes the desired domain or which mediates the desired activity *in vitro* can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a  
20 substitution or addition into the SP protein sequence. Non-classical amino acids include but are not limited to the D-isomers of the common amino acids,  $\alpha$ -amino isobutyric acid, 4-aminobutyric acid, Abu, 2-aminobutyric acid,  $\epsilon$ -Abu,  $\epsilon$ -Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid,  
25 t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine,  $\beta$ -alanine, fluoro-amino acids, designer amino acids such as  $\beta$ -methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general.

Manipulations of SP protein sequences can be made at the protein level. Also contemplated herein are SP proteins, domains thereof, derivatives or  
30 analogs or fragments thereof, which are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage

-67-

to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications can be carried out by known techniques, including but not limited to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH<sub>4</sub>, acetylation, formylation, oxidation, reduction and  
5 metabolic synthesis in the presence of tunicamycin.

In cases where natural products are suspected of having a mutation or are isolated from new species, the amino acid sequence of the SP protein isolated from the natural source, as well as those expressed *in vitro*, or from synthesized expression vectors *in vivo* or *in vitro*, can be determined from  
10 analysis of the DNA sequence, or alternatively, by direct sequencing of the isolated protein. Such analysis can be performed by manual sequencing or through use of an automated amino acid sequenator.

In particular, the protease domain of the CVSP17 can be expressed intracellularly without a signal sequence, which results in accumulation or  
15 formation of inclusion bodies containing protease domain. The inclusion bodies are isolated, denatured, solubilized and refolded protease domain, which is then activated by cleavage at the RI site.

#### Modifications

A variety of modifications of the SP proteins and domains are  
20 contemplated herein. An SP-encoding nucleic acid molecule be modified by any of numerous strategies known in the art (Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The sequences can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if  
25 desired, isolated, and ligated *in vitro*. In the production of the gene encoding a domain, derivative or analog of SP, care should be taken to ensure that the modified gene retains the original translational reading frame, uninterrupted by translational stop signals, in the gene region where the desired activity is encoded.

30 Additionally, the SP-encoding nucleic acid molecules can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or form new restriction

-68-

endonuclease sites or destroy pre-existing ones, to facilitate further *in vitro* modification. Also, as described herein muteins with primary sequence alterations, such as replacements of Cys residues and elimination of glycosylation sites are contemplated. Such mutations can be effected by any  
5 technique for mutagenesis known in the art, including, but not limited to, chemical mutagenesis and *in vitro* site-directed mutagenesis (Hutchinson et al., *J. Biol. Chem.* 253:6551-6558 (1978)), use of TAB® linkers (Pharmacia). In one embodiment, for example, an SP protein or domain thereof is modified to include a fluorescent label. In other specific embodiments, the SP protein is modified to  
10 have a heterofunctional reagent, which can be used to crosslink the members of the complex.

#### F. Screening methods

The single chain protease domains can be used in a variety of methods to identify compounds that modulate the activity thereof. For SPs that exhibit  
15 higher activity or expression in tumor cells, compounds that inhibit the proteolytic activity are of particular interest. For any SPs that are active at lower levels in tumor cells, compounds or agents that enhance the activity are potentially of interest. In all instances the identified compounds include agents that are candidate cancer treatments.

20 Several types of assays are described herein. It is understood that the protease domains can be used in other assays. The single chain protease domains exhibit catalytic activity. As such they are useful for *in vitro* screening assays, including, for example in binding assays.

The CVSP17 full length zymogens, activated enzymes, single and two  
25 chain protease domains are contemplated for use in any screening assay known to those of skill in the art, including those provided herein. Hence the following description, if directed to proteolytic assays is intended to apply to use of a single chain protease domain or a catalytically active portion thereof of any SP, including a CVSP17. Other assays, such as binding assays are provided herein,  
30 particularly for use with a CVSP17, including any variants, such as splice variants thereof.

-69-

**1. Catalytic Assays for identification of agents that modulate the protease activity of an SP protein**

Methods for identifying a modulator of the catalytic activity of an SP, particularly a single chain protease domain or catalytically active portion thereof, are provided herein. The methods can be practiced by: a) contacting the CVSP17, a full-length zymogen or activated form, and particularly a single-chain domain thereof, with a substrate of the CVSP17 in the presence of a test substance, and detecting the proteolysis of the substrate, whereby the activity of the CVSP17 is assessed, and comparing the activity to a control. For example, the control can be the activity of the CVSP17 assessed by contacting a CVSP17, including a full-length zymogen or activated form, and particularly a single-chain domain thereof, particularly a single-chain domain thereof, with a substrate of the CVSP17, and detecting the proteolysis of the substrate, whereby the activity of the CVSP17 is assessed. The results in the presence and absence of the test compounds are compared. A difference in the activity indicates that the test substance modulates the activity of the CVSP17. Modulators, such as activators, of CVSP17 activation are also contemplated; such assays are discussed below.

In one embodiment a plurality of the test substances are screened simultaneously in the above screening method. In another embodiment, the CVSP17 is isolated from a target cell as a means for then identifying agents that are potentially specific for the target cell.

In another embodiment, a test substance is a therapeutic compound, and whereby a difference of the CVSP17 activity measured in the presence and in the absence of the test substance indicates that the target cell responds to the therapeutic compound.

One method includes the steps of (a) contacting the CVSP17 polypeptide or protease domain thereof with one or a plurality of test compounds under conditions conducive to interaction between the ligand and the compounds; and (b) identifying one or more compounds in the plurality that specifically binds to the ligand.

-70-

Another method provided herein includes the steps of a) contacting a CVSP17 polypeptide or protease domain thereof with a substrate of the CVSP17 polypeptide, and detecting the proteolysis of the substrate, whereby the activity of the CVSP17 polypeptide is assessed; b) contacting the CVSP17 polypeptide with a substrate of the CVSP17 polypeptide in the presence of a test substance, and detecting the proteolysis of the substrate, whereby the activity of the CVSP17 polypeptide is assessed; and c) comparing the activity of the CVSP17 polypeptide assessed in steps a) and b), whereby the activity measured in step a) differs from the activity measured in step b) indicates that the test substance modulates the activity of the CVSP17 polypeptide.

In another embodiment, a plurality of the test substances are screened simultaneously. In comparing the activity of a CVSP17 polypeptide in the presence and absence of a test substance to assess whether the test substance is a modulator of the CVSP17 polypeptide, it is unnecessary to assay the activity in parallel, although such parallel measurement is typical. It is possible to measure the activity of the CVSP17 polypeptide at one time point and compare the measured activity to a historical value of the activity of the CVSP17 polypeptide.

For instance, one can measure the activity of the CVSP17 polypeptide in the presence of a test substance and compare with historical value of the activity of the CVSP17 polypeptide measured previously in the absence of the test substance, and *vice versa*. This can be accomplished, for example, by providing the activity of the CVSP17 polypeptide on an insert or pamphlet provided with a kit for conducting the assay. Methods for selecting substrates for a particular SP are described in the EXAMPLES, and particular proteolytic assays are described.

Combinations and kits containing the combinations optionally including instructions for performing the assays are provided. The combinations include a CVSP17 polypeptide and a substrate of the CVSP17 polypeptide to be assayed; and, optionally reagents for detecting proteolysis of the substrate. The substrates, which are can be chromogenic or fluorogenic molecules, including proteins, subject to proteolysis by a particular CVSP17 polypeptide, can be

-71-

identified empirically by testing the ability of the CVSP17 polypeptide to cleave the test substrate. Substrates that are cleaved most effectively (i.e., at the lowest concentrations and/or fastest rate or under desirable conditions), are identified.

- 5            Additionally provided herein is a kit containing the above-described combination. The kit optionally includes instructions for identifying a modulator of the activity of a CVSP17 polypeptide. Any CVSP17 polypeptide is contemplated as target for identifying modulators of the activity thereof.

## 2. Binding assays

- 10           Also provided herein are methods for identification and isolation of agents, particularly compounds that bind to CVSP17s. The assays are designed to identify agents that bind to the zymogen form, the single chain isolated protease domain (or a protein, other than a CVSP17 polypeptide, that contains the protease domain of a CVSP17 polypeptide), and to the activated form,
- 15           including the activated form derived from the full length zymogen or from polypeptide that contains the protease domain. The identified compounds are candidates or leads for identification of compounds for treatments of tumors and other disorders and diseases involving aberrant proliferation and/or angiogenesis. The CVSP17 polypeptides used in the methods include any CVSP17 polypeptide
- 20           as defined herein, including the CVSP17 single chain protease domain or proteolytically active portion thereof.

- A variety of methods are provided herein. These methods can be performed in solution or in solid phase reactions in which the CVSP17 polypeptide(s) or protease domain(s) thereof are linked, either directly or
- 25           indirectly via a linker, to a solid support. Screening assays are described in the Examples, and these assays can be used to identify candidate compounds. For purposes herein, all binding assays described above are provided for CVSP17.

- Methods for identifying an agent, such as a compound, that specifically binds to a CVSP17 single chain protease domain, a zymogen or full-length
- 30           activated CVSP17 and/or two chain protease domain thereof or other polypeptides provided herein are provided. The method can be practiced by (a) contacting the CVSP17 with one or a plurality of test agents under conditions

-72-

conducive to binding between the CVSP17 and an agent; and (b) identifying one or more agents within the plurality that specifically binds to one or more CVSP17 forms.

For example, in practicing such methods the CVSP17 polypeptide is mixed  
5 with a potential binding partner or an extract or fraction of a cell under conditions that allow the association of potential binding partners with the polypeptide. After mixing, peptides, polypeptides, proteins or other molecules that have become associated with a CVSP17 are separated from the mixture. The binding partner that bound to the CVSP17 can then be removed and further  
10 analyzed. To identify and isolate a binding partner, the entire protein, for instance the entire disclosed protein of SEQ ID Nos. 6 can be used. Alternatively, a fragment of the protein can be used.

A variety of methods can be used to obtain cell extracts or body fluids, such as blood, serum, urine, sweat, synovial fluid, CSF and other such fluids.  
15 For example, cells can be disrupted using either physical or chemical disruption methods. Examples of physical disruption methods include, but are not limited to, sonication and mechanical shearing. Examples of chemical lysis methods include, but are not limited to, detergent lysis and enzyme lysis. A skilled artisan can readily adapt methods for preparing cellular extracts in order to obtain  
20 extracts for use in the present methods.

Once an extract of a cell is prepared, the extract is mixed with the CVSP17 under conditions in which association of the protein with the binding partner can occur. A variety of conditions can be used, including conditions that resemble conditions found in the cytoplasm of a human cell. Features such as  
25 osmolarity, pH, temperature, and the concentration of cellular extract used, can be varied to optimize the association of the protein with the binding partner. Similarly, methods for isolation of molecules of interest from body fluids are known.

After mixing under appropriate conditions, the bound complex is  
30 separated from the mixture. A variety of techniques can be used to separate the mixture. For example, antibodies specific to a CVSP17 can be used to immunoprecipitate the binding partner complex. Alternatively, standard chemical

-73-

separation techniques such as chromatography and density/sediment centrifugation can be used.

After removing the non-associated cellular constituents in the extract, the binding partner can be dissociated from the complex using conventional  
5 methods. For example, dissociation can be accomplished by altering the salt concentration and/or pH of the mixture.

To aid in separating associated binding partner pairs from the mixed extract, the CVSP17 can be immobilized on a solid support. For example, the protein can be attached to a nitrocellulose matrix or acrylic beads. Attachment  
10 of the protein or a fragment thereof to a solid support aids in separating peptide/binding partner pairs from other constituents found in the extract. The identified binding partners can be either a single protein or a complex made up of two or more proteins.

Alternatively, the nucleic acid molecules encoding the single chain  
15 proteases can be used in a yeast two-hybrid system. The yeast two-hybrid system has been used to identify other protein partner pairs and can readily be adapted to employ the nucleic acid molecules herein described.

Another *in vitro* binding assay, particularly for a CVSP17, uses a mixture of a polypeptide that contains at least the catalytic domain of one of these  
20 proteins and one or more candidate binding targets or substrates. After incubating the mixture under appropriate conditions, the ability of the CVSP17 or a polypeptide fragment thereof containing the catalytic domain to bind to or interact with the candidate substrate is assessed. For cell-free binding assays, one of the components includes or is coupled to a detectable label. The label  
25 can provide for direct detection, such as radioactivity, luminescence, including fluorescence, optical or electron density, or indirect detection such as an epitope tag and an enzyme. A variety of methods can be employed to detect the label depending on the nature of the label and other assay components. For example, the label can be detected bound to the solid substrate or a portion of the bound  
30 complex containing the label can be separated from the solid substrate, and the label thereafter detected.

-74-

### 3. Detection of signal transduction

Secreted CVSPs, such as CVSP17, can be involved in signal transduction either directly by binding to or interacting with a cell surface receptor or indirectly by activating proteins, such as pro-growth factors that can initiate  
5 signal transduction. Assays for assessing signal transduction are well known to those of skill in the art, and can be adapted for use with the CVSP17 polypeptide.

Assays for identifying agents that affect or alter signal transduction mediated directly or indirectly, such as via activation of a pro-growth factor, by a  
10 CVSP17, particularly the full length or a sufficient portion to anchor the extracellular domain or a functional portion thereof of a CVSP on the surface of a cell are provided. Such assays, include, for example, transcription based assays in which modulation of a transduced signal is assessed by detecting an effect on an expression from a reporter gene (see, *e.g.*, U.S. Patent No. 5,436,128).

### 15 4. Methods for Identifying Agents that Modulate the Expression a Nucleic Acid Encoding a CVSP17

Another embodiment provides methods for identifying agents that modulate the expression of a nucleic acid encoding a CVSP17. Such assays use any available means of monitoring for changes in the expression level of the  
20 nucleic acids encoding a CVSP17.

In one assay format, cell lines that contain reporter gene fusions between the open reading frame of CVSP17 or a domain thereof, particularly the protease domain and any assayable fusion partner can be prepared. Numerous assayable fusion partners are known and readily available including the firefly luciferase  
25 gene and the gene encoding chloramphenicol acetyltransferase (Alam *et al.*, *Anal. Biochem.* 188: 245-54 (1990)). Cell lines containing the reporter gene fusions are then exposed to the agent to be tested under appropriate conditions and time. Differential expression of the reporter gene between samples exposed to the agent and control samples identifies agents which modulate the  
30 expression of a nucleic acid encoding a CVSP17.

Additional assay formats can be used to monitor the ability of the agent to modulate the expression of a nucleic acid encoding a CVSP17. For instance,

-75-

mRNA expression can be monitored directly by hybridization to the nucleic acids. Cell lines are exposed to the agent to be tested under appropriate conditions and time and total RNA or mRNA is isolated by standard procedures (see, *e.g.*, Sambrook *et al.* (1989) MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed. Cold Spring Harbor Laboratory Press). Probes to detect differences in RNA expression levels between cells exposed to the agent and control cells can be prepared from the nucleic acids. It is typical, but not necessary, to design probes which hybridize only with target nucleic acids under conditions of high stringency. Only highly complementary nucleic acid hybrids form under conditions of high stringency. Accordingly, the stringency of the assay conditions determines the amount of complementarity which should exist between two nucleic acid strands in order to form a hybrid. Stringency should be chosen to maximize the difference in stability between the probe:target hybrid and potential probe:non-target hybrids.

Probes can be designed from the nucleic acids through methods known in the art. For instance, the G + C content of the probe and the probe length can affect probe binding to its target sequence. Methods to optimize probe specificity are commonly available (see, *e.g.*, Sambrook *et al.* (1989) MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed. Cold Spring Harbor Laboratory Press); and Ausubel *et al.* (1995) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Greene Publishing Co., NY).

Hybridization conditions are modified using known methods (see, *e.g.*, Sambrook *et al.* (1989) MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed. Cold Spring Harbor Laboratory Press); and Ausubel *et al.* (1995) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Greene Publishing Co., NY), as required for each probe. Hybridization of total cellular RNA or RNA enriched for polyA RNA can be accomplished in any available format. For instance, total cellular RNA or RNA enriched for polyA RNA can be affixed to a solid support, and the solid support exposed to at least one probe comprising at least one, or part of one of the nucleic acid molecules under conditions in which the probe specifically hybridizes. Alternatively, nucleic acid fragments comprising at least one, or part of one of the sequences can be affixed to a solid support, such as a

-76-

porous glass wafer. The glass wafer can then be exposed to total cellular RNA or polyA RNA from a sample under conditions in which the affixed sequences specifically hybridize. Such glass wafers and hybridization methods are widely available, for example, those disclosed by Beattie (WO 95/11755). By

- 5 examining for the ability of a given probe to specifically hybridize to an RNA sample from an untreated cell population and from a cell population exposed to the agent, agents which up or down regulate the expression of a nucleic acid encoding the CVSP17 polypeptide, are identified.

- 10 In one format, the relative amounts of a protein between a cell population that has been exposed to the agent to be tested compared to an un-exposed control cell population can be assayed (*e.g.*, a prostate cancer cell line, a lung cancer cell line, a colon cancer cell line or a breast cancer cell line). In this format, probes, such as specific antibodies, are used to monitor the differential expression or level of activity of the protein in the different cell populations or
- 15 body fluids. Cell lines or populations or body fluids are exposed to the agent to be tested under appropriate conditions and time. Cellular lysates or body fluids can be prepared from the exposed cell line or population and from a control, unexposed cell line or population or unexposed body fluid. The cellular lysates or body fluids are then analyzed with the probe.

- 20 For example, N- and C- terminal fragments of the CVSP17 can be expressed in bacteria and used to search for proteins which bind to these fragments. Fusion proteins, such as His-tag or GST fusion to the N- or C-terminal regions of the CVSP17 can be prepared for use as a substrate. These fusion proteins can be coupled to, for example, Glutathione-Sepharose beads and
- 25 then probed with cell lysates or body fluids. Prior to lysis, the cells or body fluids can be treated with a candidate agent which can modulate a CVSP17 or proteins that interact with domains thereon. Lysate proteins binding to the fusion proteins can be resolved by SDS-PAGE, isolated and identified by protein sequencing or mass spectroscopy, as is known in the art.

- 30 Antibody probes are prepared by immunizing suitable mammalian hosts in appropriate immunization protocols using the peptides, polypeptides or proteins if they are of sufficient length (*e.g.*, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20,

-77-

25, 30, 35, 40 or more consecutive amino acids the CVSP17 polypeptide or if required to enhance immunogenicity, conjugated to suitable carriers. Methods for preparing immunogenic conjugates with carriers, such as bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), or other carrier proteins are well known in the art. In some circumstances, direct conjugation using, for example, carbodiimide reagents can be effective; in other instances linking reagents such as those supplied by Pierce Chemical Co., Rockford, IL, can be desirable to provide accessibility to the hapten. Hapten peptides can be extended at either the amino or carboxy terminus with a Cys residue or interspersed with cysteine residues, for example, to facilitate linking to a carrier. Administration of the immunogens is conducted generally by injection over a suitable time period and with use of suitable adjuvants, as is generally understood in the art. During the immunization schedule, titers of antibodies are taken to determine adequacy of antibody formation.

Anti-peptide antibodies can be generated using synthetic peptides corresponding to, for example, the carboxy terminal amino acids of the CVSP17. Synthetic peptides can be as small as 1-3 amino acids in length, generally at least 4 or more amino acid residues long. The peptides can be coupled to KLH using standard methods and can be immunized into animals, such as rabbits or ungulate. Polyclonal antibodies can then be purified, for example using Actigel beads containing the covalently bound peptide.

While the polyclonal antisera produced in this way can be satisfactory for some applications, for pharmaceutical compositions, use of monoclonal preparations are generally used. Immortalized cell lines which secrete the desired monoclonal antibodies can be prepared using the standard method of Kohler *et al.*, (*Nature* 256: 495-7 (1975)) or modifications which effect immortalization of lymphocytes or spleen cells, as is generally known. The immortalized cell lines secreting the desired antibodies are screened by immunoassay in which the antigen is the peptide hapten, polypeptide or protein. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either *in vitro* or by production *in vivo* via

-78-

ascites fluid. Of particular interest, are monoclonal antibodies that recognize the catalytic domain of an CVSP17.

Additionally, the zymogen or two-chain form of the CVSP17 can be used to make monoclonal antibodies that recognize conformation epitopes. The  
5 desired monoclonal antibodies are then recovered from the culture supernatant or from the ascites supernatant. Fragments of the monoclonals or the polyclonal antisera that contain the antigen binding portion can be used as antagonists, as well as the intact antibodies. Immunologically reactive fragments, such as the Fab, Fab', or F(ab')<sub>2</sub> fragments are often used, especially in a therapeutic  
10 context, as these fragments are generally less immunogenic than the whole immunoglobulin. Regions that bind specifically to the desired regions of receptor also can be produced in the context of chimeras with multiple species origin.

**G. Assay formats and selection of test substances that modulate at least one activity of a CVSP17 polypeptide**

15 Methods for identifying agents that modulate at least one activity of a CVSP17 are provided. The methods include phage display and other methods for assessing alterations in the activity of a CVSP17. Such methods or assays can use any means of monitoring or detecting the desired activity. A variety of formats and detection protocols are known for performing screening assays.  
20 Any such formats and protocols can be adapted for identifying modulators of CVSP17 polypeptide activities. The following includes a discussion of exemplary protocols.

**1. High throughput screening assays**

Although the above-described assay can be conducted where a single  
25 CVSP17 polypeptide is screened, and/or a single test substance is screened in one assay, the assay typically is conducted in a high throughput screening mode, *i.e.*, a plurality of the SP proteins are screened against and/or a plurality of the test substances are screened simultaneously (*See generally, High Throughput Screening: The Discovery of Bioactive Substances* (Devlin, Ed.)  
30 Marcel Dekker, 1997; Sittampalam et al., *Curr. Opin. Chem. Biol.*, 1:384-91 (1997); and Silverman et al., *Curr. Opin. Chem. Biol.*, 2:397-403 (1998)). For

-79-

example, the assay can be conducted in a multi-well (e.g., 24-, 48-, 96-, 384-, 1536-well or higher density), chip or array format.

High-throughput screening (HTS) is the process of testing a large number of diverse chemical structures against disease targets to identify "hits"

5 (Sittampalam et al., *Curr. Opin. Chem. Biol.*, 1:384-91 (1997)). Current state-of-the-art HTS operations are highly automated and computerized to handle sample preparation, assay procedures and the subsequent processing of large volumes of data.

Detection technologies employed in high-throughput screens depend on  
10 the type of biochemical pathway being investigated (Sittampalam et al., *Curr. Opin. Chem. Biol.*, 1:384-91 (1997)). These methods include, radiochemical methods, such as the scintillation proximity assays (SPA), which can be adapted to a variety of enzyme assays (Lerner et al., *J. Biomol. Screening*, 1:135-143 (1996); Baker et al., *Anal. Biochem.*, 239:20-24 (1996); Baum et al., *Anal.*  
15 *Biochem.*, 237:129-134 (1996); and Sullivan et al., *J. Biomol. Screening* 2:19-23 (1997)) and protein-protein interaction assays (Braunwalder et al., *J. Biomol. Screening* 1:23-26 (1996); Sonatore et al., *Anal. Biochem.* 240:289-297 (1996); and Chen et al., *J. Biol. Chem.* 271:25308-25315 (1996)), and non-isotopic detection methods, including but are not limited to, colorimetric and  
20 luminescence detection methods, resonance energy transfer (RET) methods, time-resolved fluorescence (HTRF) methods, cell-based fluorescence assays, such as fluorescence resonance energy transfer (FRET) procedures (see, e.g., Gonzalez et al., *Biophys. J.*, 69:1272-1280 (1995)), fluorescence polarization or anisotropy methods (see, e.g., Jameson et al., *Methods Enzymol.*  
25 246:283-300 (1995); Jolley, *J. Biomol. Screening* 1:33-38 (1996); Lynch et al., *Anal. Biochem.* 247:77-82 (1997)), fluorescence correlation spectroscopy (FCS) and other such methods.

## 2. Test Substances

Test compounds, including small molecules, antibodies, proteins, nucleic  
30 acids, peptides, natural products, mixtures of natural products, derivatives (e.g., chemical derivatives) of natural products, and libraries and collections thereof, can be screened in the above-described assays and assays described below to

-80-

identify compounds that modulate the activity of a CVSP17 polypeptide. Rational drug design methodologies that rely on computational chemistry can be used to screen and identify candidate compounds.

Test compounds (agents) that are assayed in the methods can be  
5 produced and obtained by any method known to those of skill in the art. For example, they can be randomly selected or rationally selected or designed. The agents can be, as examples, peptides, small molecules, and carbohydrates. A skilled artisan can readily recognize that there is no limit to the structural nature of the agents. The peptide agents can be prepared using standard solid phase  
10 (or solution phase) peptide synthesis methods, as is known in the art. In addition, the DNA encoding these peptides can be synthesized using commercially available oligonucleotide synthesis instrumentation and produced recombinantly using standard recombinant production systems. The production using solid phase peptide synthesis is necessitated if non-gene-encoded amino  
15 acids are to be included.

The compounds identified by the screening methods include inhibitors, including antagonists, and can be agonists. Compounds for screening include any compounds and collections of compounds available, known or that can be prepared.

#### 20 a. Selection of Compounds

Compounds can be selected for their potency and selectivity of inhibition of serine proteases, especially a CVSP17 polypeptide. As described herein, and as generally known, a target serine protease and its substrate are combined under assay conditions permitting reaction of the protease with its substrate.  
25 The assay is performed in the absence of test compound, and in the presence of increasing concentrations of the test compound. The concentration of test compound at which 50% of the serine protease activity is inhibited by the test compound is the  $IC_{50}$  value (Inhibitory Concentration) or  $EC_{50}$  (Effective Concentration) value for that compound. Within a series or group of test  
30 compounds, those having lower  $IC_{50}$  or  $EC_{50}$  values are considered more potent inhibitors of the serine protease than those compounds having higher  $IC_{50}$  or  $EC_{50}$  values. The  $IC_{50}$  measurement is often used for more simplistic assays,

-81-

whereas the  $EC_{50}$  is often used for more complicated assays, such as those employing cells.

Typically candidate compounds have an  $IC_{50}$  value of 100 nM or less as measured in an *in vitro* assay for inhibition of CVSP17 polypeptide activity. The  
5 test compounds also are evaluated for selectivity toward a serine protease. As described herein, and as generally known, a test compound is assayed for its potency toward a panel of serine proteases and other enzymes and an  $IC_{50}$  value or  $EC_{50}$  value is determined for each test compound in each assay system. A compound that demonstrates a low  $IC_{50}$  value or  $EC_{50}$  value for the target  
10 enzyme, *e.g.*, CVSP17 polypeptide, and a higher  $IC_{50}$  value or  $EC_{50}$  value for other enzymes within the test panel (*e.g.*, urokinase tissue plasminogen activator, thrombin, Factor Xa), is considered to be selective toward the target enzyme. Generally, a compound is deemed selective if its  $IC_{50}$  value or  $EC_{50}$  value in the target enzyme assay is at least 2-fold, 5-fold, 10-fold (or higher-fold)  
15 less than the next smallest  $IC_{50}$  value or  $EC_{50}$  value measured in the selectivity panel of enzymes.

Compounds are also evaluated for their activity *in vivo*. The type of assay chosen for evaluation of test compounds depends on the pathological condition to be treated or prevented by use of the compound, as well as the  
20 route of administration to be evaluated for the test compound.

For instance, to evaluate the activity of a compound to reduce tumor growth through inhibition of CVSP17 polypeptide, the procedures described by Jankun et al., *Canc. Res.* 57:559-563 (1997) to evaluate PAI-1 can be employed. Briefly, the ATCC cell lines DU145 and LnCaP are injected into SCID  
25 mice. After tumors are established, the mice are given test compound according to a dosing regime determined from the compound's *in vitro* characteristics. The Jankun *et al.* compound was administered in water. Tumor volume measurements are taken twice a week for about five weeks. A compound is deemed active if an animal to which the compound was administered exhibited  
30 decreased tumor volume, as compared to animals receiving appropriate control compounds.

-82-

Another *in vivo* experimental model designed to evaluate the effect of p-aminobenzamidine, a swine protease inhibitor, on reducing tumor volume is described by Billström et al., *Int. J. Cancer* 61:542-547 (1995).

To evaluate the ability of a compound to reduce the occurrence of, or  
5 inhibit, metastasis, the procedures described by Kobayashi *et al. Int. J. Canc.*  
57:727-733d (1994) can be employed. Briefly, a murine xenograft selected for  
high lung colonization potential is injected into C57B1/6 mice i.v. (experimental  
metastasis) or s.c. into the abdominal wall (spontaneous metastasis). Various  
concentrations of the compound to be tested can be admixed with the tumor  
10 cells in Matrigel prior to injection. Daily i.p. injections of the test compound are  
made either on days 1-6 or days 7-13 after tumor inoculation. The animals are  
sacrificed about three or four weeks after tumor inoculation, and the lung tumor  
colonies are counted. Evaluation of the resulting data permits a determination as  
to efficacy of the test compound, optimal dosing and route of administration.

15 The activity of the tested compounds toward decreasing tumor volume  
and metastasis can be evaluated in model described in Rabbani et al., *Int. J.*  
*Cancer* 63:840-845 (1995) to evaluate their inhibitor. There, Mat LyLu tumor  
cells were injected into the flank of Copenhagen rats. The animals were  
implanted with osmotic minipumps to continuously administer various doses of  
20 test compound for up to three weeks. The tumor mass and volume of  
experimental and control animals were evaluated during the experiment, as were  
metastatic growths. Evaluation of the resulting data permits a determination as  
to efficacy of the test compound, optimal dosing, and route of administration.  
Some of these authors described a related protocol in Xing et al., *Canc. Res.*  
25 57:3585-3593 (1997).

To evaluate the anti-angiogenesis activity of a compound, a rabbit cornea  
neovascularization model can be employed (see, *e.g.*, Avery *et al.* (1990) *Arch.*  
*Ophthalmol.*, 108:1474-147). Avery *et al.* describes anesthetizing New Zealand  
albino rabbits and then making a central corneal incision and forming a radial  
30 corneal pocket. A slow release prostaglandin pellet was placed in the pocket to  
induce neovascularization. Test compound was administered i.p. for five days,  
at which time the animals were sacrificed. The effect of the test compound is

-83-

evaluated by review of periodic photographs taken of the limbus, which can be used to calculate the area of neovascular response and, therefore, limbal neovascularization. A decreased area of neovascularization as compared with appropriate controls indicates the test compound was effective at decreasing or  
5 inhibiting neovascularization.

An angiogenesis model used to evaluate the effect of a test compound in preventing angiogenesis is described by Min *et al.* *Canc. Res.* 56:2428-2433 (1996). C57BL6 mice receive subcutaneous injections of a Matrigel mixture containing bFGF, as the angiogenesis-inducing agent, with and without the test  
10 compound. After five days, the animals are sacrificed and the Matrigel plugs, in which neovascularization can be visualized, are photographed. An experimental animal receiving Matrigel and an effective dose of test compound exhibits less vascularization than a control animal or an experimental animal receiving a less- or non-effective dose of compound.

15 An *in vivo* system designed to test compounds for their ability to limit the spread of primary tumors is described by Crowley *et al.*, *Proc. Natl. Acad. Sci.* 90:5021-5025 (1993). Nude mice are injected with tumor cells (PC3) engineered to express CAT (chloramphenicol acetyltransferase). Compounds to be tested for their ability to decrease tumor size and/or metastases are  
20 administered to the animals, and subsequent measurements of tumor size and/or metastatic growths are made. In addition, the level of CAT detected in various organs provides an indication of the ability of the test compound to inhibit metastasis; detection of less CAT in tissues of a treated animal versus a control animal indicates less CAT-expressing cells migrated to that tissue.

25 *In vivo* experimental models designed to evaluate the inhibitory potential of a test serine protease inhibitors, using a tumor cell line F3II known to be highly invasive (see, *e.g.*, Alonso *et al.*, *Breast Canc. Res. Treat.* 40:209-223 (1996)) can be used. Alonso describes *in vivo* studies for toxicity determination, tumor growth, invasiveness, spontaneous metastasis, experimental lung  
30 metastasis, and an angiogenesis assay.

The CAM model (chick embryo chorioallantoic membrane model), first described by L. Ossowski in 1998 (*J. Cell Biol.* 107:2437-2445 (1988)),

-84-

provides another method for evaluating the inhibitory activity of a test compound. In the CAM model, tumor cells invade through the chorioallantoic membrane. Administration of several serine protease inhibitors resulted in less or no invasion of the tumor cells through the membrane. Thus, the CAM assay  
5 is performed with CAM and tumor cells in the presence and absence of various concentrations of test compound. The invasiveness of tumor cells is measured under such conditions to provide an indication of the compound's inhibitory activity. A compound having inhibitory activity correlates with less tumor invasion.

10 The CAM model is also used in a standard assay of angiogenesis (*i.e.*, effect on formation of new blood vessels (Brooks *et al. Methods in Molecular Biology* 129:257-269 (1999)). In this model, a filter disc containing an angiogenesis inducer, such as basic fibroblast growth factor (bFGF) is placed onto the CAM. Diffusion of the cytokine into the CAM induces local  
15 angiogenesis, which can be measured in several ways such as by counting the number of blood vessel branch points within the CAM directly below the filter disc. The ability of identified compounds to inhibit cytokine-induced angiogenesis can be tested using this model. A test compound can either be added to the filter disc that contains the angiogenesis inducer, be placed directly  
20 on the membrane or be administered systemically. The extent of new blood vessel formation in the presence and/or absence of test compound can be compared using this model. The formation of fewer new blood vessels in the presence of a test compound would be indicative of anti-angiogenesis activity. A demonstration of anti-angiogenesis activity for inhibitors of a CVSP17  
25 polypeptide is indicative of a role in angiogenesis for that SP protein.

**b. Known serine protease inhibitors**

Compounds for screening can be serine protease inhibitors, which can be tested for their ability to inhibit the activity of a CVSP17. Exemplary, serine protease inhibitors for use in the screening assays, include,  
30 but are not limited to: Serine Protease Inhibitor 3 (SPI-3) (Chen, *et al. Cytokine*, 11:856-862 (1999)); Aprotinin (Iijima, R., *et al., J. Biochem. (Tokyo)* 126:912-916 (1999)); Kazal-type serine protease inhibitor-like proteins (Niimi, *et al. Eur.*

- J. Biochem.*, 266:282-292 (1999)); Kunitz-type serine protease inhibitor (Ravichandran, S., et al., *Acta Crystallogr. D. Biol. Crystallogr.*, 55:1814-1821 (1999)); Tissue factor pathway inhibitor-2/Matrix-associated serine protease inhibitor (TFPI-2/MSPI), (Liu, Y. et al. *Arch. Biochem. Biophys.* 370:112-8 (1999)); Bukunin (Cui, C.Y. et al. *J. Invest. Dermatol.* 113:182-8 (1999));
- 5 Nafmostat mesilate (Ryo, R. et al. *Vox Sang.* 76:241-6 (1999)); TPCK (Huang et al. *Oncogene* 18:3431-3439 (1999)); A synthetic cotton-bound serine protease inhibitor (Edwards et al. *Wound Repair Regen.* 7:106-18 (1999)); FUT-175 (Sawada, M. et al. *Stroke* 30:644-50 (1999)); Combination of serine
- 10 protease inhibitor FUT-0175 and thromboxane synthetase inhibitor OKY-046 (Kaminogo et al. *Neurol. Med. Chir. (Tokyo)* 38:704-8; discussion 708-9 (1998)); The rat serine protease inhibitor 2.1 gene (LeCam, A., et al., *Biochem. Biophys. Res. Commun.*, 253:311-4 (1998)); A new intracellular serine protease inhibitor expressed in the rat pituitary gland complexes with granzyme B (Hill et
- 15 al. *FEBS Lett.* 440:361-4 (1998)); 3,4-Dichloroisocoumarin (Hammed et al. *Proc. Soc. Exp. Biol. Med.*, 219:132-7 (1998)); LEX032 (Bains et al. *Eur. J. Pharmacol.* 356:67-72 (1998)); N-tosyl-L-phenylalanine chloromethyl ketone (Dryjanski et al. *Biochemistry* 37:14151-6 (1998)); Mouse gene for the serine
- 20 protease inhibitor neuroserpin (P112) (Berger et al. *Gene*, 214:25-33 (1998)); Rat serine protease inhibitor 2.3 gene (Paul et al. *Eur. J. Biochem.* 254:538-46 (1998)); Ecotin (Yang et al. *J. Mol. Biol.* 279:945-57 (1998)); A 14 kDa plant-related serine protease inhibitor (Roch et al. *Dev. Comp. Immunol.* 22(1):1-12 (1998)); Matrix-associated serine protease inhibitor TFPI-2/33 kDa MSPI (Rao et
- 25 al. *Int. J. Cancer* 76:749-56 (1998)); ONO-3403 (Hiwasa et al. *Cancer Lett.* 126:221-5 (1998)); Bdellastasin (Moser et al. *Eur. J. Biochem.* 253:212-20 (1998)); Bikunin (Xu et al. *J. Mol. Biol.* 276:955-66 (1998)); Nafamostat mesilate (Mellgren et al. *Thromb. Haemost.* 79:342-7 (1998)); The growth
- hormone dependent serine protease inhibitor, Spi 2.1 (Maake et al. *Endocrinology* 138:5630-6 (1997)); Growth factor activator inhibitor type 2, a
- 30 Kunitz-type serine protease inhibitor (Kawaguchi et al. *J. Biol. Chem.*, 272:27558-64 (1997)); Heat-stable serine protease inhibitor protein from ovaries of the desert locust, *Schistocerca gregaria* (Hamdaoui et al. *Biochem. Biophys.*

- Res. Commun.* 238:357-60 (1997)); Human placental Hepatocyte growth factor activator inhibitor, a Kunitz-type serine protease inhibitor (Shimomura *et al.* *J. Biol. Chem.* 272:6370-6 (1997)); FUT-187, oral serine protease inhibitor (Shiozaki *et al.* *Gan To Kagaku Ryoho*, 23(14): 1971-9 (1996)); Extracellular
- 5 matrix-associated serine protease inhibitors (Mr 33,000, 31,000, and 27,000 (Rao, C.N., *et al.*, *Arch. Biochem. Biophys.*, 335:82-92 (1996)); An irreversible isocoumarin serine protease inhibitor (Palencia, D.D., *et al.*, *Biol. Reprod.*, 55:536-42 (1996)); 4-(2-aminoethyl)-benzenesulfonyl fluoride (AEBSF) (Nakabo *et al.* *J. Leukoc. Biol.* 60:328-36 (1996)); Neuroserpin (Osterwalder, T., *et al.*,
- 10 *EMBO J.* 15:2944-53 (1996)); Human serine protease inhibitor alpha-1-antitrypsin (Forney *et al.* *J. Parasitol.* 82:496-502 (1996)); Rat serine protease inhibitor 2.3 (Simar-Blanchet, A.E., *et al.*, *Eur. J. Biochem.*, 236:638-48 (1996)); Gebaxate mesilate (parodi, F., *et al.*, *J. Cardiothorac. Vasc. Anesth.* 10:235-7 (1996)); Recombinant serine protease inhibitor, CPTI II (Stankiewicz, M., *et al.*,
- 15 *Acta Biochim. Pol.*, 43(3):525-9 (1996)); A cysteine-rich serine protease inhibitor (Guamerin II) (Kim, D.R., *et al.*, *J. Enzym. Inhib.*, 10:81-91 (1996)); Diisopropylfluorophosphate (Lundqvist, H., *et al.*, *Inflamm. Res.*, 44(12):510-7 (1995)); Nexin 1 (Yu, D.W., *et al.*, *J. Cell Sci.*, 108(Pt 12):3867-74 (1995)); LEX032 (Scalia, R., *et al.*, *Shock*, 4(4):251-6 (1995)); Protease nexin I
- 20 (Houenou, L.J., *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 92(3):895-9 (1995)); Chymase-directed serine protease inhibitor (Woodard S.L., *et al.*, *J. Immunol.*, 153(11):5016-25 (1994)); N-alpha-tosyl-L-lysyl-chloromethyl ketone (TLCK) (Bourinbaier, A.S., *et al.*, *Cell Immunol.*, 155(1):230-6 (1994)); Smpi56 (Ghendler, Y., *et al.*, *Exp. Parasitol.*, 78(2):121-31 (1994)); Schistosoma
- 25 haematobium serine protease inhibitor (Blanton, R.E., *et al.*, *Mol. Biochem. Parasitol.*, 63(1):1-11 (1994)); Spi-1 (Warren, W.C., *et al.*, *Mol. Cell Endocrinol.*, 98(1):27-32 (1993)); TAME (Jessop, J.J., *et al.*, *Inflammation*, 17(5):613-31 (1993)); Antithrombin III (Kalaria, R.N., *et al.*, *Am. J. Pathol.*, 143(3):886-93 (1993)); FOY-305 (Ohkoshi, M., *et al.*, *Anticancer Res.*, 13(4):963-6 (1993));
- 30 Camostat mesilate (Senda, S., *et al.*, *Intern. Med.*, 32(4):350-4 (1993)); Pigment epithelium-derived factor (Steele, F.R., *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 90(4):1526-30 (1993)); Antistatin (Holstein, T.W., *et al.*, *FEBS Lett.*,

-87-

- 309(3):288-92 (1992)); the vaccinia virus K2L gene encodes a serine protease inhibitor (Zhou, J., et al., *Virology*, 189(2):678-86 (1992)); Bowman-Birk serine-protease inhibitor (Werner, M.H., et al., *J. Mol. Biol.*, 225(3):873-89 (1992); FUT-175 (Yanamoto, H., et al., *Neurosurgery*, 30(3):358-63 (1992)); FUT-175; 5 (Yanamoto, H., et al., *Neurosurgery*, 30(3):351-6, discussion 356-7 (1992)); PAI-I (Yreadwell, B.V., et al., *J. Orthop. Res.*, 9(3):309-16 (1991)); 3,4-Dichloroisocoumarin (Rusbridge, N.M., et al., *FEBS Lett.*, 268(1):133-6 (1990)); Alpha 1-antichymotrypsin (Lindmark, B.E., et al., *Am. Rev. Respir. Dis.*, 141(4 Pt 1):884-8 (1990)); P-toluenesulfonyl-L-arginine methyl ester (TAME) (Scuderi, 10 P., *J. Immunol.*, 143(1):168-73 (1989)); Alpha 1-antichymotrypsin (Abraham, C.R., et al., *Cell*, 52(4):487-501 (1988)); Contrapsin (Modha, J., et al., *Parasitology*, 96 (Pt 1):99-109 (1988)); Alpha 2-antiplasmin (Holmes, W.E., et al., *J. Biol. Chem.*, 262(4):1659-64 (1987)); 3,4-dichloroisocoumarin (Harper, J.W., et al., *Biochemistry*, 24(8):1831-41 (1985)); Diisopropylfluorophosphate 15 (Tsutsui, K., et al., *Biochem. Biophys. Res. Commun.*, 123(1):271-7 (1984)); Gabexate mesilate (Hesse, B., et al., *Pharmacol. Res. Commun.*, 16(7):637-45 (1984)); Phenyl methyl sulfonyl fluoride (Dufer, J., et al., *Scand. J. Haematol.*, 32(1):25-32 (1984)); Protease inhibitor Cl-2 (McPhalen, C.A., et al., *J. Mol. Biol.*, 168(2):445-7 (1983)); Phenylmethylsulfonyl fluoride (Sekar V., et al., 20 *Biochem. Biophys. Res. Commun.*, 89(2):474-8 (1979)); PGE1 (Feinstein, M.D., et al., *Prostaglandine*, 14(6):1075-93 (1977)).

#### c. Combinatorial libraries and other libraries

- The source of compounds for the screening assays, can be libraries, including, but are not limited to, combinatorial libraries. Methods for 25 synthesizing combinatorial libraries and characteristics of such combinatorial libraries are known in the art (See generally, *Combinatorial Libraries: Synthesis, Screening and Application Potential* (Cortese Ed.) Walter de Gruyter, Inc., 1995; Tietze and Lieb, *Curr. Opin. Chem. Biol.*, 2(3):363-71 (1998); Lam, *Anticancer Drug Des.*, 12(3):145-67 (1997); Blaney and Martin, *Curr. Opin. Chem. Biol.*, 30 1(1):54-9 (1997); and Schultz and Schultz, *Biotechnol. Prog.*, 12(6):729-43 (1996)).

Methods and strategies for generating diverse libraries, primarily peptide- and nucleotide-based oligomer libraries, have been developed using molecular biology methods and/or simultaneous chemical synthesis methodologies (see, e.g., Dower et al., *Annu. Rep. Med. Chem.*, 26:271-280 (1991); Fodor et al., *Science*, 251:767-773 (1991); Jung et al., *Angew. Chem. Ind. Ed. Engl.*, 31:367-383 (1992); Zuckerman et al., *Proc. Natl. Acad. Sci. USA*, 89:4505-4509 (1992); Scott et al., *Science*, 249:386-390 (1990); Devlin et al., *Science*, 249:404-406 (1990); Cwirla et al., *Proc. Natl. Acad. Sci. USA*, 87:6378-6382 (1990); and Gallop et al., *J. Medicinal Chemistry*, 37:1233-1251 (1994)). The resulting combinatorial libraries potentially contain millions of compounds that can be screened to identify compounds that exhibit a selected activity.

The libraries fall into roughly three categories: fusion-protein-displayed peptide libraries in which random peptides or proteins are presented on the surface of phage particles or proteins expressed from plasmids; support-bound synthetic chemical libraries in which individual compounds or mixtures of compounds are presented on insoluble matrices, such as resin beads (see, e.g., Lam et al., *Nature*, 354:82-84 (1991)) and cotton supports (see, e.g., Eichler et al., *Biochemistry* 32:11035-11041 (1993)); and methods in which the compounds are used in solution (see, e.g., Houghten et al., *Nature*, 354:84-86 (1991); Houghten et al., *BioTechniques*, 313:412-421 (1992); and Scott et al., *Curr. Opin. Biotechnol.*, 5:40-48 (1994)). There are numerous examples of synthetic peptide and oligonucleotide combinatorial libraries and there are many methods for producing libraries that contain non-peptidic small organic molecules. Such libraries can be based on a basis set of monomers that are combined to form mixtures of diverse organic molecules or that can be combined to form a library based upon a selected pharmacophore monomer.

Either a random or a deterministic combinatorial library can be screened by the presently disclosed and/or claimed screening methods. In either of these two libraries, each unit of the library is isolated and/or immobilized on a solid support. In the deterministic library, one knows *a priori* a particular unit's location on each solid support. In a random library, the location of a particular unit is not known *a priori* although each site still contains a single unique unit.

- Many methods for preparing libraries are known to those of skill in this art (see, e.g., Geysen et al., *Proc. Natl. Acad. Sci. USA*, 81:3998-4002 (1984), Houghten et al., *Proc. Natl. Acad. Sci. USA*, 81:5131-5135 (1985)).
- Combinatorial library generated by any techniques known to those of skill in the
- 5 art are contemplated (see, e.g., Table 1 of Schultz and Schultz, *Biotechnol. Prog.*, 12(6):729-43 (1996)) for screening; Bartel et al., *Science*, 261:1411-1418 (1993); Baumbach et al. *BioPharm*, (Can):24-35 (1992); Bock et al. *Nature*, 355:564-566 (1992); Borman, S., Combinatorial chemists focus on small molecules molecular recognition, and automation, *Chem. Eng. News*,
- 10 2(12):29 (1996); Boublik, et al., Eukaryotic Virus Display: Engineering the Major Surface Glycoproteins of the Autographa California Nuclear Polyhedrosis Virus (ACNPV) for the Presentation of Foreign Proteins on the Virus Surface, *Bio/Technology*, 13:1079-1084 (1995); Brenner, et al., Encoded Combinatorial Chemistry, *Proc. Natl. Acad. Sci. U.S.A.*, 89:5381-5383 (1992); Caflisch, et al.,
- 15 Computational Combinatorial Chemistry for *De Novo* Ligand Design: Review and Assessment, *Perspect. Drug Discovery Des.*, 3:51-84 (1995); Cheng, et al., Sequence-Selective Peptide Binding with a Peptido-A,B-*trans*-steroidal Receptor Selected from an Encoded Combinatorial Library, *J. Am. Chem. Soc.*, 118:1813-1814 (1996); Chu, et al., Affinity Capillary Electrophoresis to Identify the
- 20 Peptide in A Peptide Library that Binds Most Tightly to Vancomycin, *J. Org. Chem.*, 58:648-652 (1993); Clackson, et al., Making Antibody Fragments Using Phage Display Libraries, *Nature*, 352:624-628 (1991); Combs, et al., Protein Structure-Based Combinatorial Chemistry: Discovery of Non-Peptide Binding Elements to Src SH3 Domain, *J. Am. Chem. Soc.*, 118:287-288 (1996); Cwirla,
- 25 et al., Peptides On Phage: A Vast Library of Peptides for Identifying Ligands, *Proc. Natl. Acad. Sci. U.S.A.*, 87:6378-6382 (1990); Ecker, et al., Combinatorial Drug Discovery: Which Method will Produce the Greatest Value, *Bio/Technology*, 13:351-360 (1995); Ellington, et al., *In Vitro* Selection of RNA Molecules That Bind Specific Ligands, *Nature*, 346:818-822 (1990); Ellman,
- 30 J.A., Variants of Benzodiazepines, *J. Am. Chem. Soc.*, 114:10997 (1992); Erickson, et al., *The Proteins*; Neurath, H., Hill, R.L., Eds.: Academic: New York, 1976; pp. 255-257; Felici, et al., *J. Mol. Biol.*, 222:301-310 (1991); Fodor, et

-90-

- al., Light-Directed, Spatially Addressable Parallel Chemical Synthesis, *Science*, 251:767-773 (1991); Francisco, et al., Transport and Anchoring of Beta-Lactamase to the External Surface of *E. Coli.*, *Proc. Natl. Acad. Sci. U.S.A.*, 89:2713-2717 (1992); Georgiou, et al., Practical Applications of Engineering
- 5 Gram-Negative Bacterial Cell Surfaces, *TIBTECH*, 11:6-10 (1993); Geysen, et al., Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid, *Proc. Natl. Acad. Sci. U.S.A.*, 81:3998-4002 (1984); Glaser, et al., Antibody Engineering by Condon-Based Mutagenesis in a Filamentous Phage Vector System, *J. Immunol.*, 149:3903-3913 (1992); Gram, et al., In
- 10 vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library, *Proc. Natl. Acad. Sci.*, 89:3576-3580 (1992); Han, et al., Liquid-Phase Combinatorial Synthesis, *Proc. Natl. Acad. Sci. U.S.A.*, 92:6419-6423 (1995); Hoogenboom, et al., Multi-Subunit Proteins on the Surface of Filamentous Phage: Methodologies for Displaying Antibody (Fab)
- 15 Heavy and Light Chains, *Nucleic Acids Res.*, 19:4133-4137 (1991); Houghten, et al., General Method for the Rapid Solid-Phase Synthesis of Large Numbers of Peptides: Specificity of Antigen-Antibody Interaction at the Level of Individual Amino Acids, *Proc. Natl. Acad. Sci. U.S.A.*, 82:5131-5135 (1985); Houghten, et al., The Use of Synthetic Peptide Combinatorial Libraries for the Determination
- 20 of Peptide Ligands in Radio-Receptor Assays-Opioid-Peptides, *Bioorg. Med. Chem. Lett.*, 3:405-412 (1993); Houghten, et al., Generation and Use of Synthetic Peptide Combinatorial Libraries for Basic Research and Drug Discovery, *Nature*, 354:84-86 (1991); Huang, et al., Discovery of New Ligand Binding Pathways in Myoglobin by Random Mutagenesis, *Nature Struct. Biol.*, 1:226-229
- 25 (1994); Huse, et al., Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire In Phage Lambda, *Science*, 246:1275-1281 (1989); Janda, K.D., New Strategies for the Design of Catalytic Antibodies, *Biotechnol. Prog.*, 6:178-181 (1990); Jung, et al., Multiple Peptide Synthesis Methods and Their Applications, *Angew. Chem. Int. Ed. Engl.*, 31:367-486 (1992); Kang, et
- 30 al., Linkage of Recognition and Replication Functions By Assembling Combinatorial Antibody Fab Libraries Along Phage Surfaces, *Proc. Natl. Acad. Sci. U.S.A.*, 88:4363-4366 (1991a); Kang, et al., Antibody Redesign by Chain

- Shuffling from Random Combinatorial Immunoglobulin Libraries, *Proc. Natl. Acad. Sci. U.S.A.*, 88:11120-11123 (1991b); Kay, et al., An M13 Phage Library Displaying Random 38-Amino-Acid-Peptides as a Source of Novel Sequences with Affinity to Selected Targets Genes, *Gene*, 128:59-65 (1993); Lam, et al., A  
5 new type of synthetic peptide library for identifying ligand-binding activity, *Nature*, 354:82-84 (1991) (published errata in *Nature*, 358:434 (1992) and *Nature*, 360:768 (1992); Lebl, et al., One Bead One Structure Combinatorial Libraries, *Biopolymers (Pept. Sci.)*, 37:177-198 (1995); Lerner, et al., Antibodies without Immunization, *Science*, 258:1313-1314 (1992); Li, et al., Minimization  
10 of a Polypeptide Hormone, *Science*, 270:1657-1660 (1995); Light, et al., Display of Dimeric Bacterial Alkaline Phosphatase on the Major Coat Protein of Filamentous Bacteriophage, *Bioorg. Med. Chem. Lett.*, 3:1073-1079 (1992); Little, et al., Bacterial Surface Presentation of Proteins and Peptides: An Alternative to Phage Technology, *Trends Biotechnol.*, 11:3-5 (1993); Marks, et  
15 al., By-Passing Immunization. Human Antibodies from V-Gene Libraries Displayed on Phage, *J. Mol. Biol.*, 222:581-597 (1991); Matthews, et al., Substrate Phage: Selection of Protease Substrates by Monovalent Phage Display, *Science*, 260:1113-1117 (1993); McCafferty, et al., Phage Enzymes: Expression and Affinity Chromatography of Functional Alkaline Phosphatase on the Surface  
20 of Bacteriophage, *Protein Eng.*, 4:955-961 (1991); Menger, et al., Phosphatase Catalysis Developed Via Combinatorial Organic Chemistry, *J. Org. Chem.*, 60:6666-6667 (1995); Nicolaou, et al., *Angew. Chem. Int. Ed. Engl.*, 34:2289-2291 (1995); Oldenburg, et al., Peptide Ligands for A Sugar-Binding Protein Isolated from a Random Peptide Library, *Proc. Natl. Acad. Sci. U.S.A.*, 89:5393-  
25 5397 (1992); Parmley, et al., Antibody-Selectable Filamentous fd Phage Vectors: Affinity Purification of Target Genes, *Genes*, 73:305-318 (1988); Pinilla, et al., Synthetic Peptide Combinatorial Libraries (SPCLS)--Identification of the Antigenic Determinant of Beta-Endorphin Recognized by Monoclonal Antibody-3E7, *Gene*, 128:71-76 (1993); Pinilla, et al., Review of the Utility of Soluble Combinatorial  
30 Libraries, *Biopolymers*, 37:221-240 (1995); Pistor, et al., Expression of Viral Hemagglutinin On the Surface of *E. Coli.*, *Klin. Wochenschr.*, 66:110-116 (1989); Pollack, et al., Selective Chemical Catalysis by an Antibody, *Science*,

- 234:1570-1572 (1986); Rigler, et al., Fluorescence Correlations, Single Molecule Detection and Large Number Screening: Applications in Biotechnology, *J. Biotechnol.*, 41:177-186 (1995); Sarvetnick, et al., Increasing the Chemical Potential of the Germ-Line Antibody Repertoire, *Proc. Natl. Acad. Sci. U.S.A.*, 90:4008-4011 (1993); Sastry, et al., Cloning of the Immunological Repertoire in *Escherichia Coli* for Generation of Monoclonal Catalytic Antibodies: Construction of a Heavy Chain Variable Region-Specific cDNA Library, *Proc. Natl. Acad. Sci. U.S.A.*, 86:5728-5732 (1989); Scott, et al., Searching for Peptide Ligands with an Epitope Library, *Science*, 249:386-390 (1990); Sears, et al., Engineering Enzymes for Bioorganic Synthesis: Peptide Bond Formation, *Biotechnol. Prog.*, 12:423-433 (1996); Simon, et. al., Peptides: A Modular Approach to Drug Discovery, *Proc. Natl. Acad. Sci. U.S.A.*, 89:9367-9371 (1992); Still, et al., Discovery of Sequence-Selective Peptide Binding by Synthetic Receptors Using Encoded Combinatorial Libraries, *Acc. Chem. Res.*, 29:155-163 (1996); Thompson, et al., Synthesis and Applications of Small Molecule Libraries, *Chem. Rev.*, 96:555-600 (1996); Tramontano, et al., Catalytic Antibodies, *Science*, 234:1566-1570 (1986); Wrighton, et al., Small Peptides as Potent Mimetics of the Protein Hormone Erythropoietin, *Science*, 273:458-464 (1996); York, et al., Combinatorial mutagenesis of the reactive site region in plasminogen activator inhibitor I, *J. Biol. Chem.*, 266:8595-8600 (1991); Zebedee, et al., Human Combinatorial Antibody Libraries to Hepatitis B Surface Antigen, *Proc. Natl. Acad. Sci. U.S.A.*, 89:3175-3179 (1992); Zuckerman, et al., Identification of Highest-Affinity Ligands by Affinity Selection from Equimolar Peptide Mixtures Generated by Robotic Synthesis, *Proc. Natl. Acad. Sci. U.S.A.*, 89:4505-4509 (1992).

For example, peptides that bind to a CVSP17 polypeptide or a protease domain of an SP protein can be identified using phage display libraries. In an exemplary embodiment, this method can include a) contacting phage from a phage library with the CVSP17 polypeptide or a protease domain thereof; (b) isolating phage that bind to the protein; and (c) determining the identity of at least one peptide coded by the isolated phage to identify a peptide that binds to a CVSP17 polypeptide.

-93-

**H. Modulators of the activity of CVSP17 polypeptides**

Provided herein are compounds, identified by screening or produced using the CVSP17 polypeptide or protease domain in other screening methods, that modulate the activity of a CVSP17. These compounds act by directly interacting  
5 with the CVSP17 polypeptide or by altering transcription or translation thereof. Such molecules include, but are not limited to, antibodies that specifically react with a CVSP17 polypeptide, particularly with the protease domain thereof, antisense nucleic acids or double-stranded RNA (dsRNA) such as RNAi, including those that contain modified nucleic acids, that alter expression of the CVSP17  
10 polypeptide, peptide mimetics and other such compounds.

**1. Antibodies**

Antibodies, including polyclonal and monoclonal antibodies, that specifically bind to the CVSP17 polypeptide provided herein, particularly to the single chain protease domains thereof or the activated forms of the full-length or  
15 protease domain or the zymogen form, are provided.

Generally, the antibody is a monoclonal antibody, and typically the antibody specifically binds to the protease domain of the CVSP17 polypeptide. Provided are antibodies that specifically bind to any domain of CVSP17, and antibodies that specifically bind to two chain forms thereof. Also provided are  
20 antibodies that specifically bind to the active site of the zymogen and activated forms. Neutralizing antibodies are also provided. Also provided are antibodies that specifically bind to the leucine zipper-containing regions of CVSP17 are provided, particularly those that specifically bind to amino acids 397-427 of SEQ  
ID No. 6 or corresponding regions in other CVSP17 polypeptides).

25 The CVSP17 polypeptide and domains, fragments, homologs and derivatives thereof can be used as immunogens to generate antibodies that specifically bind CVSP17 polypeptides and portions thereof. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library. In a specific embodiment, antibodies  
30 to human CVSP17 polypeptide are produced. In another embodiment, complexes formed from fragments of CVSP17 polypeptide, which fragments

-94-

contain the serine protease domain, are used as immunogens for antibody production.

Various procedures known in the art can be used for the production of polyclonal antibodies to CVSP17 polypeptide, its domains, derivatives, fragments or analogs. For production of the antibody, various host animals can be immunized by injection with the native CVSP17 polypeptide or a synthetic version, or a derivative of the foregoing, such as a cross-linked CVSP17 polypeptide. Such host animals include but are not limited to rabbits, mice, rats, etc. Various adjuvants can be used to increase the immunological response, depending on the host species, and include but are not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, and potentially useful human adjuvants such as bacille Calmette-Guerin (BCG) and corynebacterium parvum.

For preparation of monoclonal antibodies directed towards a CVSP17 polypeptide or domains, derivatives, fragments or analogs thereof, any technique that provides for the production of antibody molecules by continuous cell lines in culture can be used. Such techniques include but are not restricted to the hybridoma technique originally developed by Kohler and Milstein (*Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., *Immunology Today* 4:72 (1983)), and the EBV hybridoma technique to produce human monoclonal antibodies (Cole et al., *in Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985)). For example, immortalized cell lines that secrete the desired monoclonal antibodies are. The immortalized cell lines secreting the desired antibodies are screened by immunoassays in which the antigen is the peptide hapten, polypeptide or protein. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either *in vitro* or by production *in vivo* via ascites fluid.

Monoclonal antibodies can be produced by other methods, such as in germ-free animals utilizing recent technology (PCT/US90/02545). Human antibodies can be used and can be obtained by using human hybridomas (Cote

et al., *Proc. Natl. Acad. Sci. USA* 80:2026-2030 (1983)), or by transforming human B cells with EBV virus *in vitro* (Cole et al., in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985)). Techniques developed for the production of "chimeric antibodies" (Morrison et al., *Proc. Natl. Acad. Sci. USA* 81:6851-6855 (1984); Neuberger et al., *Nature* 312:604-608 (1984); Takeda et al., *Nature* 314:452-454 (1985)) by splicing the genes from a mouse antibody molecule specific for the CVSP17 polypeptide together with genes from a human antibody molecule of appropriate biological activity can be used.

Techniques described for the production of single chain antibodies (U.S. patent 4,946,778) can be adapted to produce CVSP17 polypeptide-specific single chain antibodies. An additional embodiment uses the techniques described for the construction of Fab expression libraries (Huse et al., *Science* 246:1275-1281 (1989)) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for CVSP17 polypeptide or domains, derivatives, or analogs thereof. Non-human antibodies can be "humanized" by known methods (see, e.g., U.S. Patent No. 5,225,539).

Antibody fragments that specifically bind to CVSP17 polypeptide or epitopes thereof can be generated by techniques known in the art. For example, such fragments include but are not limited to: the F(ab')<sub>2</sub> fragment, which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments that can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragment; the Fab fragments that can be generated by treating the antibody molecule with papain and a reducing agent; and Fv fragments.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, e.g., ELISA (enzyme-linked immunosorbent assay). To select antibodies specific to a particular domain of the CVSP17 polypeptide one can assay generated hybridomas for a product that binds to the fragment of the CVSP17 polypeptide that contains such a domain

The foregoing antibodies can be used in methods known in the art relating to the localization and/or quantitation of CVSP17 polypeptide proteins, e.g., for imaging these proteins, measuring levels thereof in appropriate physiological samples, in, for example, diagnostic methods. In another

-96-

embodiment, anti-CVSP17 polypeptide antibodies, or fragments thereof, containing the binding domain are used as therapeutic agents.

## 2. Peptides, Polypeptides and Peptide Mimetics

Provided herein are methods for identifying molecules that bind to and  
5 modulate the activity of SP proteins. Included among molecules that bind to  
SPs, particularly the single chain protease domain or catalytically active  
fragments thereof, are peptides, polypeptides and peptide mimetics, including  
cyclic peptides. Peptide mimetics are molecules or compounds that mimic the  
necessary molecular conformation of a ligand or polypeptide for specific binding  
10 to a target molecule such as a CVSP17 polypeptide. In an exemplary  
embodiment, the peptides, polypeptides or peptide mimetics bind to the protease  
domain of the CVSP17 polypeptide. Such peptides and peptide mimetics include  
those of antibodies that specifically bind to a CVSP17 polypeptide and, typically,  
bind to the protease domain of a CVSP17 polypeptide. The peptides,  
15 polypeptides and peptide mimetics identified by methods provided herein can be  
agonists or antagonists of CVSP17 polypeptides.

Such peptides and peptide mimetics are useful for diagnosing, treating,  
preventing, and screening for a disease or disorder associated with CVSP17  
polypeptide activity in a mammal. In addition, the peptides and peptide mimetics  
20 are useful for identifying, isolating, and purifying molecules or compounds that  
modulate the activity of a CVSP17 polypeptide, or specifically bind to a CVSP17  
polypeptide, generally the protease domain of a CVSP17 polypeptide. Low  
molecular weight peptides and peptide mimetics can have strong binding  
properties to a target molecule, *e.g.*, a CVSP17 polypeptide and/or the protease  
25 domain of a CVSP17 polypeptide.

Peptides, polypeptides and peptide mimetics that bind to CVSP17  
polypeptides as described herein can be administered to mammals, including  
humans, to modulate CVSP17 polypeptide activity. Thus, methods for  
therapeutic treatment and prevention of neoplastic diseases comprise  
30 administering a peptide, polypeptides or peptide mimetic compound in an amount  
sufficient to modulate such activity are provided. Thus, also provided herein are  
methods for treating a subject having such a disease or disorder in which a

-97-

peptide, polypeptides or peptide mimetic compound is administered to the subject in a therapeutically effective dose or amount.

Compositions containing the peptides, polypeptides or peptide mimetics can be administered for prophylactic and/or therapeutic treatments. In

5 therapeutic applications, compositions can be administered to a patient already suffering from a disease, as described above, in an amount sufficient to cure or at least partially arrest the symptoms of the disease and its complications. Amounts effective for this use will depend on the severity of the disease and the weight and general state of the patient and can be empirically determined.

10 In prophylactic applications, compositions containing the peptides, polypeptides and peptide mimetics are administered to a patient susceptible to or otherwise at risk of a particular disease. Such an amount is defined to be a "prophylactically effective dose". In this use, the precise amounts again depend on the patient's state of health and weight. Accordingly, the peptides, 15 polypeptides and peptide mimetics that bind to a CVSP17 polypeptide can be used to prepare pharmaceutical compositions containing, as an active ingredient, at least one of the peptides or peptide mimetics in association with a pharmaceutical carrier or diluent. The compounds can be administered, for example, by oral, pulmonary, parental (intramuscular, intraperitoneal, intravenous 20 (IV) or subcutaneous injection), inhalation (via a fine powder formulation), transdermal, nasal, vaginal, rectal, or sublingual routes of administration and can be formulated in dosage forms appropriate for each route of administration (see, e.g., International PCT application Nos. WO 93/25221 and WO 94/17784; and European Patent Application 613,683).

25 Peptides, polypeptides and peptide mimetics that bind to CVSP17 polypeptides are useful *in vitro* as tools for understanding the biological role of CVSP17 polypeptides, including the evaluation of the many factors thought to influence, and be influenced by, the production of CVSP17 polypeptide. Such peptides, polypeptides and peptide mimetics also are useful in the development 30 of other compounds that bind to and modulate the activity of a CVSP17 polypeptide, because such compounds provide important information on the

-98-

relationship between structure and activity that should facilitate such development.

The peptides, polypeptides and peptide mimetics are also useful as competitive binders in assays to screen for new CVSP17 polypeptides or  
5 CVSP17 polypeptide agonists. In such assay embodiments, the compounds can be used without modification or can be modified in a variety of ways; for example, by labeling, such as covalently or non-covalently joining a moiety which directly or indirectly provides a detectable signal. In any of these assays, the materials thereto can be labeled either directly or indirectly. Possibilities for  
10 direct labeling include label groups such as: radiolabels such as <sup>125</sup>I enzymes (U.S. Pat. No. 3,645,090) such as peroxidase and alkaline phosphatase, and fluorescent labels (U.S. Pat. No. 3,940,475) capable of monitoring the change in fluorescence intensity, wavelength shift, or fluorescence polarization. Possibilities for indirect labeling include biotinylation of one constituent followed  
15 by binding to avidin coupled to one of the above label groups. The compounds can also include spacers or linkers in cases where the compounds are to be attached to a solid support.

Moreover, based on their ability to bind to a CVSP17 polypeptide, the peptides, polypeptides and peptide mimetics can be used as reagents for  
20 detecting CVSP17 polypeptides in living cells, fixed cells, in biological fluids, in tissue homogenates and in purified, natural biological materials. For example, by labeling such peptides, polypeptides and peptide mimetics, cells having CVSP17 polypeptides can be identified. In addition, based on their ability to bind a CVSP17 polypeptide, the peptides, polypeptides and peptide mimetics can  
25 be used in *in situ* staining, FACS (fluorescence-activated cell sorting), Western blotting, ELISA and other analytical protocols. Based on their ability to bind to a CVSP17 polypeptide, the peptides, polypeptides and peptide mimetics can be used in purification of CVSP17 polypeptides or in purifying cells expressing the CVSP17 polypeptide, *e.g.*, a polypeptide encoding the protease domain of a  
30 CVSP17 polypeptide.

The peptides, polypeptides and peptide mimetics can also be used as commercial reagents for various medical research and diagnostic uses. The

-99-

activity of the peptides and peptide mimetics can be evaluated either *in vitro* or *in vivo* in one of the numerous models described in McDonald (1992) *Am. J. of Pediatric Hematology/Oncology*, 14:8-21.

### 3. Peptide, polypeptides and peptide mimetic therapy

5        Peptide analogs are commonly used in the pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compounds are termed "peptide mimetics" or "peptidomimetics" (Luthman *et al.*, *A Textbook of Drug Design and Development*, 14:386-406, 2nd Ed., Harwood Academic Publishers (1996);  
10    Joachim Grante (1994) *Angew. Chem. Int. Ed. Engl.*, 33:1699-1720; Fauchere (1986) *J. Adv. Drug Res.*, 15:29; Veber and Freidinger (1985) *TINS*, p. 392; and Evans *et al.* (1987) *J. Med. Chem.* 30:1229). Peptide mimetics that are structurally similar to therapeutically useful peptides can be used to produce an equivalent or enhanced therapeutic or prophylactic effect. Preparation of  
15    peptidomimetics and structures thereof are known to those of skill in this art.

Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) can be used to generate more stable peptides. In addition, constrained peptides containing a consensus sequence or a substantially identical consensus  
20    sequence variation can be generated by methods known in the art (Rizo *et al.* (1992) *An. Rev. Biochem.*, 61:387, incorporated herein by reference); for example, by adding internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

Those skilled in the art appreciate that modifications can be made to the  
25    peptides and mimetics without deleteriously effecting the biological or functional activity of the peptide. Further, the skilled artisan would know how to design non-peptide structures in three dimensional terms, that mimic the peptides that bind to a target molecule, *e.g.*, a CVSP17 polypeptide or, generally, the protease domain of CVSP17 polypeptides (see, *e.g.*, Eck and Sprang (1989) *J. Biol.*  
30    *Chem.*, 26: 17605-18795).

When used for diagnostic purposes, the peptides and peptide mimetics can be labeled with a detectable label and, accordingly, the peptides and peptide

-100-

mimetics without such a label can serve as intermediates in the preparation of labeled peptides and peptide mimetics. Detectable labels can be molecules or compounds, which when covalently attached to the peptides and peptide mimetics, permit detection of the peptide and peptide mimetics *in vivo*, for example, in a patient to whom the peptide or peptide mimetic has been administered, or *in vitro*, e.g., in a sample or cells. Suitable detectable labels are well known in the art and include, by way of example, radioisotopes, fluorescent labels (e.g., fluorescein), and the like. The particular detectable label employed is not critical and is selected to be detectable at non-toxic levels. Selection of the such labels is well within the skill of the art.

Covalent attachment of a detectable label to the peptide or peptide mimetic is accomplished by conventional methods well known in the art. For example, when the  $^{125}\text{I}$  radioisotope is employed as the detectable label, covalent attachment of  $^{125}\text{I}$  to the peptide or the peptide mimetic can be achieved by incorporating the amino acid tyrosine into the peptide or peptide mimetic and then iodinating the peptide (see, e.g., Weaner *et al.* (1994) *Synthesis and Applications of Isotopically Labelled Compounds*, pp. 137-140). If tyrosine is not present in the peptide or peptide mimetic, incorporation of tyrosine to the N or C terminus of the peptide or peptide mimetic can be achieved by well known chemistry. Likewise,  $^{32}\text{P}$  can be incorporated onto the peptide or peptide mimetic as a phosphate moiety through, for example, a hydroxyl group on the peptide or peptide mimetic using conventional chemistry.

Labeling of peptidomimetics usually involves covalent attachment of one or more labels, directly or through a spacer (e.g., an amide group), to non-interfering position(s) on the peptidomimetic that are predicted by quantitative structure-activity data and/or molecular modeling. Such non-interfering positions generally are positions that do not form direct contacts with the macromolecules(s) to which the peptidomimetic binds to produce the therapeutic effect. Derivatization (e.g., labeling) of peptidomimetics should not substantially interfere with the desired biological or pharmacological activity of the peptidomimetic.

-101-

Peptides, polypeptides and peptide mimetics that can bind to a CVSP17 polypeptide or the protease domain of CVSP17 polypeptides and/or modulate the activity thereof, or exhibit CVSP17 protease activity, can be used for treatment of neoplastic disease. The peptides, polypeptides and peptide mimetics can be delivered, *in vivo* or *ex vivo*, to the cells of a subject in need of treatment. Further, peptides which have CVSP17 polypeptide activity can be delivered, *in vivo* or *ex vivo*, to cells which carry mutant or missing alleles encoding the CVSP17 polypeptide gene. Any of the techniques described herein or known to the skilled artisan can be used for preparation and *in vivo* or *ex vivo* delivery of such peptides, polypeptides and peptide mimetics that are substantially free of other human proteins. For example, the peptides, polypeptides and peptide mimetics can be readily prepared by expression in a microorganism or synthesis *in vitro*.

The peptides or peptide mimetics can be introduced into cells, *in vivo* or *ex vivo*, by microinjection or by use of liposomes, for example. Alternatively, the peptides, polypeptides or peptide mimetics can be taken up by cells, *in vivo* or *ex vivo*, actively or by diffusion. In addition, extracellular application of the peptide, polypeptides or peptide mimetic can be sufficient to effect treatment of a neoplastic disease. Other molecules, such as drugs or organic compounds, that: 1) bind to a CVSP17 polypeptide or protease domain thereof; or 2) have a similar function or activity to an CVSP17 polypeptide or protease domain thereof, can be used in methods for treatment.

#### 4. Rational drug design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides or peptides of interest or of small molecules or peptide mimetics with which they interact (*e.g.*, agonists and antagonists) in order to fashion drugs which are, *e.g.*, more active or stable forms thereof; or which, for example, enhance or interfere with the function of a polypeptide *in vivo* (*e.g.*, a CVSP17 polypeptide). In one approach, one first determines the three-dimensional structure of a protein of interest (*e.g.*, a CVSP17 polypeptide or polypeptide having a protease domain) or, for example, of a CVSP17 polypeptide-ligand complex, by X-ray crystallography, by computer modeling or

-102-

most typically, by a combination of approaches (see, *e.g.*, Erickson *et al.* 1990). Also, useful information regarding the structure of a polypeptide can be gained by modeling based on the structure of homologous proteins. In addition, peptides can be analyzed by an alanine scan. In this technique, an amino acid  
5 residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

Also, a polypeptide or peptide that binds to a CVSP17 polypeptide or, generally, the protease domain of a CVSP17 polypeptide, can be selected by a  
10 functional assay, and then the crystal structure of this polypeptide or peptide can be determined. The polypeptide can be, for example, an antibody specific for a CVSP17 polypeptide and/or the protease domain of a CVSP17 polypeptide. This approach can yield a pharmacophore upon which subsequent drug design can be based. Further, it is possible to bypass the crystallography altogether by  
15 generating anti-idiotypic polypeptides or peptides, (anti-ids) to a functional, pharmacologically active polypeptide or peptide that binds to a CVSP17 polypeptide or protease domain of a CVSP17 polypeptide. As a mirror image of a mirror image, the binding site of the anti-ids is expected to be an analog of the original target molecule, *e.g.*, a CVSP17 polypeptide or polypeptide having a  
20 CVSP17 polypeptide. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacophore.

Thus, one can design drugs which have, *e.g.*, improved activity or stability or which act as modulators (*e.g.*, inhibitors, agonists, antagonists) of  
25 CVSP17 polypeptide activity, and are useful in the methods, particularly the methods for diagnosis, treatment, prevention, and screening of a neoplastic disease. By virtue of the availability of cloned CVSP17 polypeptide sequences, sufficient amounts of the CVSP17 polypeptide can be made available to perform such analytical studies as X-ray crystallography. In addition, the knowledge of  
30 the amino acid sequence of a CVSP17 polypeptide or the protease domain thereof, *e.g.*, the protease domain encoded by the nucleotide sequence of SEQ

-103-

ID No. 6, can provide guidance on computer modeling techniques in place of, or in addition to, X-ray crystallography.

**Methods of identifying peptides and peptide mimetics that bind to CVSP17 polypeptides**

5       Peptides having a binding affinity to the CVSP17 polypeptide polypeptides provided herein (*e.g.*, a CVSP17 polypeptide or a polypeptide having a protease domain of a CVSP17 polypeptide) can be readily identified, for example, by random peptide diversity generating systems coupled with an affinity enrichment process. Specifically, random peptide diversity generating systems include the "peptides on plasmids" system (see, *e.g.*, U.S. Patent Nos. 10 5,270,170 and 5,338,665); the "peptides on phage" system (see, *e.g.*, U.S. Patent No. 6,121,238 and Cwirla, *et al.* (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:6378-6382); the "polysome system;" the "encoded synthetic library (ESL)" system; and the "very large scale immobilized polymer synthesis" system (see, 15 *e.g.*, U.S. Patent No. 6,121,238; and Dower *et al.* (1991) *An. Rep. Med. Chem.* 26:271-280

For example, using the procedures described above, random peptides can generally be designed to have a defined number of amino acid residues in length (*e.g.*, 12). To generate the collection of oligonucleotides encoding the random 20 peptides, the codon motif (NNK)x, where N is nucleotide A, C, G, or T (equimolar; depending on the methodology employed, other nucleotides can be employed), K is G or T (equimolar), and x is an integer corresponding to the number of amino acids in the peptide (*e.g.*, 12) can be used to specify any one of the 32 possible codons resulting from the NNK motif: 1 for each of 12 amino 25 acids, 2 for each of 5 amino acids, 3 for each of 3 amino acids, and only one of the three stop codons. Thus, the NNK motif encodes all of the amino acids, encodes only one stop codon, and reduces codon bias.

The random peptides can be presented, for example, either on the surface of a phage particle, as part of a fusion protein containing either the pIII or the 30 pVIII coat protein of a phage fd derivative (peptides on phage) or as a fusion protein with the LacI peptide fusion protein bound to a plasmid (peptides on plasmids). The phage or plasmids, including the DNA encoding the peptides, can

-104-

be identified and isolated by an affinity enrichment process using immobilized CVSP17 polypeptide having a protease domain. The affinity enrichment process, sometimes called "panning," typically involves multiple rounds of incubating the phage, plasmids, or polysomes with the immobilized CVSP17 polypeptide, collecting the phage, plasmids, or polysomes that bind to the CVSP17 polypeptide (along with the accompanying DNA or mRNA), and producing more of the phage or plasmids (along with the accompanying LacI-peptide fusion protein) collected.

#### Characteristics of peptides and peptide mimetics

Among the peptides, polypeptides and peptide mimetics for therapeutic application are those of having molecular weights from about 250 to about 8,000 daltons. If such peptides are oligomerized, dimerized and/or derivatized with a hydrophilic polymer (*e.g.*, to increase the affinity and/or activity of the compounds), the molecular weights of such peptides can be substantially greater and can range anywhere from about 500 to about 120,000 daltons, generally from about 8,000 to about 80,000 daltons. Such peptides can contain 9 or more amino acids that are naturally occurring or synthetic (non-naturally occurring) amino acids. One skilled in the art can determine the affinity and molecular weight of the peptides and peptide mimetics suitable for therapeutic and/or diagnostic purposes (*e.g.*, see Dower *et al.*, U.S. Patent No. 6,121,238).

#### 5. Methods of preparing peptides and peptide mimetics

Peptides and peptide mimetics can be designed, using a variety of methods, such as, for example, the "encoded synthetic library" or "very large scale immobilized polymer synthesis" systems (see, *e.g.*, U.S. Patent Nos. 5,925,525 and 5,902,723). Using the "encoded synthetic library" or "very large scale immobilized polymer synthesis" systems, the minimum size of a peptide with an activity of interest, such as binding to a CVSP17, can be determined. In addition all peptides that form the group of peptides that differ from the desired motif (or the minimum size of that motif) in one, two, or more residues can be prepared. This collection of peptides then can be screened for an ability to bind to a target molecule, *e.g.*, a CVSP17 polypeptide or, generally, the protease domain of a CVSP17 polypeptide. This immobilized polymer synthesis system or

-105-

other peptide synthesis methods can also be used to synthesize truncation analogs and deletion analogs and combinations of truncation and deletion analogs of the peptide compounds.

Peptides that bind to CVSP17 polypeptides can be prepared by classical methods known in the art; for example, by using standard solid phase techniques. The standard methods include exclusive solid phase synthesis, partial solid phase synthesis methods, fragment condensation, classical solution synthesis, and even by recombinant DNA technology (see, *e.g.*, Merrifield (1963) *J. Am. Chem. Soc.*, 85:2149, incorporated herein by reference.)

These procedures can also be used to synthesize peptides in which amino acids other than the 20 naturally occurring, genetically encoded amino acids are substituted at one, two, or more positions of the peptide. For instance, naphthylalanine can be substituted for tryptophan, facilitating synthesis. Other synthetic amino acids that can be substituted into the peptides include L-hydroxypropyl, L-3, 4-dihydroxy-phenylalanyl, D amino acids such as L-D-hydroxylysyl and D-D-methylalanyl, L- $\alpha$ -methylalanyl,  $\beta$  amino acids, and isoquinolyl. D amino acids and non-naturally occurring synthetic amino acids can also be incorporated into the peptides (see, *e.g.*, Roberts *et al.* (1983) *Unusual Amino/Acids in Peptide Synthesis*, 5(6):341-449).

The peptides can also be modified by phosphorylation (see, *e.g.*, W. Bannwarth *et al.* (1996) *Biorganic and Medicinal Chemistry Letters*, 6(17):2141-2146), and other methods for making peptide derivatives (see, *e.g.*, Hruby *et al.* (1990) *Biochem. J.*, 268(2):249-262). Thus, peptide compounds also serve as a basis to prepare peptide mimetics with similar biological activity.

Those of skill in the art recognize that a variety of techniques are available for constructing peptide mimetics with the same or similar desired biological activity as the corresponding peptide compound but with more favorable activity than the peptide with respect to solubility, stability, and susceptibility to hydrolysis and proteolysis (see, *e.g.*, Morgan *et al.* (1989) *An. Rep. Med. Chem.*, 24:243-252). Methods for preparing peptide mimetics modified at the N-terminal amino group, the C-terminal carboxyl group, and/or

-106-

changing one or more of the amido linkages in the peptide to a non-amido linkage are known to those of skill in the art.

Amino terminus modifications include, but are not limited to, alkylating, acetylating and adding a carbobenzoyl group, forming a succinimide group (see, 5 *e.g.*, Murray *et al.* (1995) *Burger's Medicinal Chemistry and Drug Discovery*, 5th ed., Vol. 1, Manfred E. Wolf, ed., John Wiley and Sons, Inc.). C-terminal modifications include mimetics wherein the C-terminal carboxyl group is replaced by an ester, an amide or modifications to form a cyclic peptide.

In addition to N-terminal and C-terminal modifications, the peptide 10 compounds, including peptide mimetics, advantageously can be modified with or covalently coupled to one or more of a variety of hydrophilic polymers. It has been found that when peptide compounds are derivatized with a hydrophilic polymer, their solubility and circulation half-lives can be increased and their immunogenicity is masked, with little, if any, diminishment in their binding 15 activity. Suitable nonproteinaceous polymers include, but are not limited to, polyalkylethers as exemplified by polyethylene glycol and polypropylene glycol, polylactic acid, polyglycolic acid, polyoxyalkenes, polyvinylalcohol, polyvinylpyrrolidone, cellulose and cellulose derivatives, dextran and dextran derivatives. Generally, such hydrophilic polymers have an average molecular 20 weight ranging from about 500 to about 100,000 daltons, including from about 2,000 to about 40,000 daltons and, from about 5,000 to about 20,000 daltons. The hydrophilic polymers also can have an average molecular weights of about 5,000 daltons, 10,000 daltons and 20,000 daltons. The peptide compounds can be dimerized and each of the dimeric subunits can be covalently attached to 25 a hydrophilic polymer. The peptide compounds can be PEGylated, i.e., covalently attached to polyethylene glycol (PEG).

Methods for derivatizing peptide compounds or for coupling peptides to such polymers have been described (see, *e.g.*, Zallipsky (1995) *Bioconjugate Chem.*, 6:150-165; Monfardini *et al.* (1995) *Bioconjugate Chem.*, 6:62-69; U.S. 30 Pat. No. 4,640,835; U.S. Pat. No. 4,496,689; U.S. Pat. No. 4,301,144; U.S. Pat. No. 4,670,417; U.S. Pat. No. 4,791,192; U.S. Pat. No. 4,179,337 and WO 95/34326, all of which are incorporated by reference in their entirety herein).

-107-

Other methods for making peptide derivatives are described, for example, in Hruby *et al.* (1990), *Biochem J.*, 268(2):249-262, which is incorporated herein by reference. Thus, the peptide compounds also serve as structural models for non-peptidic compounds with similar biological activity. Those of skill in the art recognize that a variety of techniques are available for constructing compounds with the same or similar desired biological activity as a particular peptide compound but with more favorable activity with respect to solubility, stability, and susceptibility to hydrolysis and proteolysis (see, *e.g.*, Morgan *et al.* (1989) *An. Rep. Med. Chem.*, 24:243-252, incorporated herein by reference). These techniques include replacing the peptide backbone with a backbone composed of phosphonates, amidates, carbamates, sulfonamides, secondary amines, and N-methylamino acids.

Peptide compounds can exist in a cyclized form with an intramolecular disulfide bond between the thiol groups of the cysteines. Alternatively, an intermolecular disulfide bond between the thiol groups of the cysteines can be produced to yield a dimeric (or higher oligomeric) compound. One or more of the cysteine residues can also be substituted with a homocysteine.

#### **I. Conjugates**

A conjugate, containing: a) a single chain protease domain (or proteolytically active portion thereof) of a CVSP17 polypeptide or a full length zymogen, activated form thereof, or two or single chain protease domain thereof; and b) a targeting agent linked to the CVSP17 polypeptide directly or via a linker, wherein the agent facilitates: i) affinity isolation or purification of the conjugate; ii) attachment of the conjugate to a surface; iii) detection of the conjugate; or iv) targeted delivery to a selected tissue or cell, is provided herein. The conjugate can be a chemical conjugate or a fusion protein mixture thereof.

The targeting agent can be a protein or peptide fragment, such as a tissue specific or tumor specific monoclonal antibody or growth factor or fragment thereof linked either directly or via a linker to a CVSP17 polypeptide or a protease domain thereof. The targeting agent can also be a protein or peptide fragment that contains a protein binding sequence, a nucleic acid binding sequence, a lipid binding sequence, a polysaccharide binding sequence, or a

-108-

metal binding sequence, or a linker for attachment to a solid support. In a particular embodiment, the conjugate contains a) the CVSP17 or portion thereof, as described herein; and b) a targeting agent linked to the CVSP17 polypeptide directly or via a linker.

5           Conjugates, such as fusion proteins and chemical conjugates, of the CVSP17 polypeptide with a protein or peptide fragment (or plurality thereof) that functions, for example, to facilitate affinity isolation or purification of the CVSP17 polypeptide domain, attachment of the CVSP17 polypeptide domain to a surface, or detection of the CVSP17 polypeptide domain are provided. The  
10   conjugates can be produced by chemical conjugation, such as via thiol linkages, and can be produced by recombinant means as fusion proteins. In the fusion protein, the peptide or fragment thereof is linked to either the N-terminus or C-terminus of the CVSP17 polypeptide domain. In chemical conjugates the peptide or fragment thereof can be linked anywhere that conjugation can be effected,  
15   and there can be a plurality of such peptides or fragments linked to a single CVSP17 polypeptide domain or to a plurality thereof.

          The targeting agent is for *in vitro* or *in vivo* delivery to a cell or tissue, and includes agents such as cell or tissue-specific antibodies, growth factors and other factors that bind to moieties expressed on specific cells; and other cell or  
20   tissue specific agents that promote directed delivery of a linked protein. The targeting agent can be one that specifically delivers the CVSP17 polypeptide to selected cells by interaction with a cell surface protein and internalization of conjugate or CVSP17 polypeptide portion thereof.

          These conjugates are used in a variety of methods and are particularly  
25   suited for use in methods of activation of prodrugs, such as prodrugs that upon cleavage by the particular CVSP17, which is localized at or near the targeted cell or tissue, are cytotoxic. The prodrugs are administered prior to, or simultaneously with, or subsequently to the conjugate. Upon delivery to the targeted cells, the protease activates the prodrug, which then exhibits a  
30   therapeutic effect, such as a cytotoxic effect.

-109-

## **1. Conjugation**

Conjugates with linked CVSP17 polypeptides and/or domains thereof can be prepared either by chemical conjugation, recombinant DNA technology, or by combinations of recombinant expression and chemical conjugation. The CVSP17  
5 polypeptide domains and the targeting agent can be linked in any orientation and more than one targeting agents and/or CVSP17 polypeptide domains can be present in a conjugate.

### **a. Fusion proteins**

Fusion proteins are provided herein. A fusion protein contains: a) one or  
10 a plurality of domains of an CVSP17 polypeptide and b) a targeting agent. The fusion proteins are generally produced by recombinant expression of nucleic acids that encode the fusion protein.

### **b. Chemical conjugation**

To effect chemical conjugation herein, the CVSP17 polypeptide domain is  
15 linked via one or more selected linkers or directly to the targeting agent. Chemical conjugation must be used if the targeted agent is other than a peptide or protein, such a nucleic acid or a non-peptide drug. Any means known to those of skill in the art for chemically conjugating selected moieties can be used.

## **2. Linkers**

20 Linkers for can be included in the conjugates. The conjugates can include one or more linkers between the CVSP17 polypeptide portion and the targeting agent. Additionally, linkers are used for facilitating or enhancing immobilization of a CVSP17 polypeptide or portion thereof on a solid support, such as a microtiter plate, silicon or silicon-coated chip, glass or plastic support, such as  
25 for high throughput solid phase screening protocols. Any linker known to those of skill in the art for preparation of conjugates can be used herein. These linkers are typically used in the preparation of chemical conjugates; peptide linkers can be incorporated into fusion proteins.

Linkers can be any moiety suitable to associate a domain of CVSP17  
30 polypeptide and a targeting agent. Such linkers and linkages include, but are not limited to, peptidic linkages, amino acid and peptide linkages, typically containing between one and about 60 amino acids, more generally between about 10 and

-110-

- 30 amino acids, chemical linkers, such as heterobifunctional cleavable cross-linkers, including but are not limited to, N-succinimidyl (4-iodoacetyl)-aminobenzoate, sulfosuccinimidyl (4-iodoacetyl)-aminobenzoate, 4-succinimidyl-oxycarbonyl- $\alpha$ -(2-pyridyldithio)toluene, sulfosuccinimidyl-6-[ $\alpha$ -methyl- $\alpha$ -
- 5 (pyridyldithiol)-toluamido] hexanoate, N-succinimidyl-3-(2-pyridyldithio) - propionate, succinimidyl 6[3-(2-pyridyldithio)-propionamido] hexanoate, sulfosuccinimidyl 6[3-(2-pyridyldithio)-propionamido] hexanoate, 3-(2-pyridyldithio)-propionyl hydrazide, Ellman's reagent, dichlorotriazinic acid, and S-(2-thiopyridyl)-L-cysteine. Other linkers include, but are not limited to peptides and
- 10 other moieties that reduce steric hindrance between the domain of CVSP17 polypeptide and the targeting agent, intracellular enzyme substrates, linkers that increase the flexibility of the conjugate, linkers that increase the solubility of the conjugate, linkers that increase the serum stability of the conjugate, photocleavable linkers and acid cleavable linkers.
- 15 Other exemplary linkers and linkages that are suitable for chemically linked conjugates include, but are not limited to, disulfide bonds, thioether bonds, hindered disulfide bonds, and covalent bonds between free reactive groups, such as amine and thiol groups. These bonds are produced using heterobifunctional reagents to produce reactive thiol groups on one or both of
- 20 the polypeptides and then reacting the thiol groups on one polypeptide with reactive thiol groups or amine groups to which reactive maleimido groups or thiol groups can be attached on the other. Other linkers include, acid cleavable linkers, such as bismaleimideoxy propane, acid labile-transferrin conjugates and adipic acid dihydrazide, that would be cleaved in more acidic intracellular
- 25 compartments; cross linkers that are cleaved upon exposure to UV or visible light and linkers, such as the various domains, such as C<sub>H</sub>1, C<sub>H</sub>2, and C<sub>H</sub>3, from the constant region of human IgG<sub>1</sub> (see, Batra *et al. Molecular Immunol.*, 30:379-386 (1993)). In some embodiments, several linkers can be included in order to take advantage of desired properties of each linker.
- 30 Chemical linkers and peptide linkers can be inserted by covalently coupling the linker to the domain of CVSP17 polypeptide and the targeting agent. The heterobifunctional agents, described below, can be used to effect

-111-

such covalent coupling. Peptide linkers can also be linked by expressing DNA encoding the linker and therapeutic agent (TA), linker and targeted agent, or linker, targeted agent and therapeutic agent (TA) as a fusion protein. Flexible linkers and linkers that increase solubility of the conjugates are contemplated for use, either alone or with other linkers are also contemplated herein.

a) **Acid cleavable, photocleavable and heat sensitive linkers**

Acid cleavable linkers, photocleavable and heat sensitive linkers can also be used, particularly where it can be necessary to cleave the domain of CVSP17 polypeptide to permit it to be more readily accessible to reaction. Acid cleavable linkers include, but are not limited to, bismaleimideoxy propane; and adipic acid dihydrazide linkers (see, *e.g.*, Fattom *et al.* (1992) *Infection & Immun.* 60:584-589) and acid labile transferrin conjugates that contain a sufficient portion of transferrin to permit entry into the intracellular transferrin cycling pathway (see, *e.g.*, Welhöner *et al.* (1991) *J. Biol. Chem.* 266:4309-4314).

Photocleavable linkers are linkers that are cleaved upon exposure to light (see, *e.g.*, Goldmacher *et al.* (1992) *Bioconj. Chem.* 3:104-107, which linkers are herein incorporated by reference), thereby releasing the targeted agent upon exposure to light. Photocleavable linkers that are cleaved upon exposure to light are known (see, *e.g.*, Hazum *et al.* (1981) in *Pept., Proc. Eur. Pept. Symp.*, 16th, Brunfeldt, K (Ed), pp. 105-110, which describes the use of a nitrobenzyl group as a photocleavable protective group for cysteine; Yen *et al.* (1989) *Makromol. Chem* 190:69-82, which describes water soluble photocleavable copolymers, including hydroxypropylmethacrylamide copolymer, glycine copolymer, fluorescein copolymer and methylrhodamine copolymer; Goldmacher *et al.* (1992) *Bioconj. Chem.* 3:104-107, which describes a cross-linker and reagent that undergoes photolytic degradation upon exposure to near UV light (350 nm); and Senter *et al.* (1985) *Photochem. Photobiol* 42:231-237, which describes nitrobenzyloxycarbonyl chloride cross linking reagents that produce photocleavable linkages), thereby releasing the targeted agent upon exposure to light. Such linkers would have particular use in treating dermatological or ophthalmic conditions that can be exposed to light using fiber optics. After administration of the conjugate, the eye or skin or other body part

-112-

can be exposed to light, resulting in release of the targeted moiety from the conjugate. Such photocleavable linkers are useful in connection with diagnostic protocols in which it is desirable to remove the targeting agent to permit rapid clearance from the body of the animal.

5                   **b) Other linkers for chemical conjugation**

Other linkers, include trityl linkers, particularly, derivatized trityl groups to generate a genus of conjugates that provide for release of therapeutic agents at various degrees of acidity or alkalinity. The flexibility thus afforded by the ability to preselect the pH range at  
10 which the therapeutic agent is released allows selection of a linker based on the known physiological differences between tissues in need of delivery of a therapeutic agent (see, e.g., U.S. Patent No. 5,612,474). For example, the acidity of tumor tissues appears to be lower than that of normal tissues.

**c) Peptide linkers**

15           The linker moieties can be peptides. Peptide linkers can be employed in fusion proteins and also in chemically linked conjugates. The peptide typically has from about 2 to about 60 amino acid residues, for example from about 5 to about 40, or from about 10 to about 30 amino acid residues. The length selected depends upon factors, such as the use for which the linker is included.

20           Peptide linkers are advantageous when the targeting agent is proteinaceous. For example, the linker moiety can be a flexible spacer amino acid sequence, such as those known in single-chain antibody research. Examples of such known linker moieties include, but are not limited to, peptides, such as  $(\text{Gly}_m\text{Ser})_n$  and  $(\text{Ser}_m\text{Gly})_n$ , in which n is 1 to 6, including 1 to  
25 4 and 2 to 4, and m is 1 to 6, including 1 to 4, and 2 to 4, enzyme cleavable linkers and others.

Additional linking moieties are described, for example, in Huston *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 85:5879-5883, 1988; Whitlow, M., *et al.*, *Protein Engineering* 6:989-995, 1993; Newton *et al.*, *Biochemistry* 35:545-553, 1996;  
30 A. J. Cumber *et al.*, *Bioconj. Chem.* 3:397-401, 1992; Ladurner *et al.*, *J. Mol. Biol.* 273:330-337, 1997; and U.S. Patent. No. 4,894,443. In some

-113-

embodiments, several linkers can be included in order to take advantage of desired properties of each linker.

### 3. Targeting agents

Any agent that facilitates detection, immobilization, or purification of the conjugate is contemplated for use herein. For chemical conjugates any moiety that has such properties is contemplated; for fusion proteins, the targeting agent is a protein, peptide or fragment thereof that is sufficient to effect the targeting activity. Contemplated targeting agents include those that deliver the CVSP17 polypeptide or portion thereof to selected cells and tissues. Such agents include tumor specific monoclonal antibodies and portions thereof, growth factors, such as FGF, EGF, PDGF, VEGF, cytokines, including chemokines, and other such agents.

### 4. Nucleic acids, plasmids and cells

Isolated nucleic acid fragments encoding fusion proteins are provided. The nucleic acid fragment that encodes the fusion protein includes: a) nucleic acid encoding a protease domain of a CVSP17 polypeptide; and b) nucleic acid encoding a protein, peptide or effective fragment thereof that facilitates: i) affinity isolation or purification of the fusion protein; ii) attachment of the fusion protein to a surface; or iii) detection of the fusion protein. Generally, the nucleic acid is DNA.

Plasmids for replication and vectors for expression that contain the above nucleic acid fragments are also provided. Cells containing the plasmids and vectors are also provided. The cells can be any suitable host including, but are not limited to, bacterial cells, yeast cells, fungal cells, plant cells, insect cell and animal cells. The nucleic acids, plasmids, and cells containing the plasmids can be prepared according to methods known in the art including any described herein.

Also provided are methods for producing the above fusion proteins. An exemplary method includes the steps of growing cells (*i.e.*, culturing the cells so that they proliferate) containing a plasmid encoding the fusion protein under conditions whereby the fusion protein is expressed by the cell, and recovering the expressed fusion protein. Methods for expressing and recovering

-114-

recombinant proteins are well known in the art (*See generally, Current Protocols in Molecular Biology* (1998) § 16, John Wiley & Sons, Inc.) and such methods can be used for expressing and recovering the expressed fusion proteins.

The recovered fusion proteins can be isolated or purified by methods known in the art such as centrifugation, filtration, chromatograph, electrophoresis, immunoprecipitation, etc., or by a combination thereof (*See generally, Current Protocols in Molecular Biology* (1998) § 10, John Wiley & Sons, Inc.). Generally the recovered fusion protein is isolated or purified through affinity binding between the protein or peptide fragment of the fusion protein and an affinity binding moiety. As discussed in the above sections regarding the construction of the fusion proteins, any affinity binding pairs can be constructed and used in the isolation or purification of the fusion proteins. For example, the affinity binding pairs can be protein binding sequences/protein, DNA binding sequences/DNA sequences, RNA binding sequences/RNA sequences, lipid binding sequences/lipid, polysaccharide binding sequences/polysaccharide, or metal binding sequences/metal.

#### **5. Immobilization and supports or substrates therefor**

In certain embodiments, where the targeting agents are designed for linkage to surfaces, the CVSP17 polypeptide can be attached by linkage such as ionic or covalent, non-covalent or other chemical interaction, to a surface of a support or matrix material. Immobilization can be effected directly or via a linker. The CVSP17 polypeptide can be immobilized on any suitable support, including, but are not limited to, silicon chips, and other supports described herein and known to those of skill in the art. A plurality of CVSP17 polypeptide or protease domains thereof can be attached to a support, such as an array (*i.e.*, a pattern of two or more) of conjugates on the surface of a silicon chip or other chip for use in high throughput protocols and formats.

It is also noted that the domains of the CVSP17 polypeptide can be linked directly to the surface or via a linker without a targeting agent linked thereto. Hence chips containing arrays of the domains of the CVSP17 polypeptide are also provided.

-115-

The matrix material or solid supports contemplated herein are generally any of the insoluble materials known to those of skill in the art to immobilize ligands and other molecules, and are those that are used in many chemical syntheses and separations. Such supports are used, for example, in affinity  
5 chromatography, in the immobilization of biologically active materials, and during chemical syntheses of biomolecules, including proteins, amino acids and other organic molecules and polymers. The preparation of and use of supports is well known to those of skill in this art; there are many such materials and preparations thereof known. For example, naturally-occurring support materials,  
10 such as agarose and cellulose, can be isolated from their respective sources, and processed according to known protocols, and synthetic materials can be prepared in accord with known protocols.

The supports are typically insoluble materials that are solid, porous, deformable, or hard, and have any required structure and geometry, including,  
15 but not limited to: beads, pellets, disks, capillaries, hollow fibers, needles, solid fibers, random shapes, thin films and membranes. Thus, the item can be fabricated from the matrix material or combined with it, such as by coating all or part of the surface or impregnating particles.

Typically, when the matrix is particulate, the particles are at least about  
20 10-2000  $\mu\text{m}$ , but can be smaller or larger, depending upon the selected application. Selection of the matrices is governed, at least in part, by their physical and chemical properties, such as solubility, functional groups, mechanical stability, surface area swelling propensity, hydrophobic or hydrophilic properties and intended use.

25 If necessary, the support matrix material can be treated to contain an appropriate reactive moiety. In some cases, the support matrix material already containing the reactive moiety can be obtained commercially. The support matrix material containing the reactive moiety can thereby serve as the matrix support upon which molecules are linked. Materials containing reactive surface  
30 moieties such as amino silane linkages, hydroxyl linkages or carboxysilane linkages can be produced by well established surface chemistry techniques involving silanization reactions, or the like. Examples of these materials are

-116-

those having surface silicon oxide moieties, covalently linked to gamma-amino-propylsilane, and other organic moieties; N-[3-(triethoxysilyl)propyl]phthalamic acid; and bis-(2-hydroxyethyl)aminopropyltriethoxysilane. Exemplary of readily available materials containing amino group reactive functionalities, include, but  
5 are not limited to, para-aminophenyltriethoxysilane. Also derivatized polystyrenes and other such polymers are well known and readily available to those of skill in this art (*e.g.*, the Tentagel® Resins are available with a multitude of functional groups, and are sold by Rapp Polymere, Tubingen, Germany; see, U.S. Patent No. 4,908,405 and U.S. Patent No. 5,292,814; *see, also* Butz et al.,  
10 *Peptide Res.*, 7:20-23 (1994); and Kleine et al., *Immunobiol.*, 190:53-66 (1994)).

These matrix materials include any material that can act as a support matrix for attachment of the molecules of interest. Such materials are known to those of skill in this art, and include those that are used as a support matrix.  
15 These materials include, but are not limited to, inorganics, natural polymers, and synthetic polymers, including, but are not limited to: cellulose, cellulose derivatives, acrylic resins, glass, silica gels, polystyrene, gelatin, polyvinyl pyrrolidone, co-polymers of vinyl and acrylamide, polystyrene cross-linked with divinylbenzene and others (*see*, Merrifield, *Biochemistry*, 3:1385-1390 (1964)),  
20 polyacrylamides, latex gels, polystyrene, dextran, polyacrylamides, rubber, silicon, plastics, nitrocellulose, celluloses, natural sponges. Of particular interest herein, are highly porous glasses (*see, e.g.*, U.S. Patent No. 4,244,721) and others prepared by mixing a borosilicate, alcohol and water.

Synthetic supports include, but are not limited to: acrylamides, dextran-  
25 derivatives and dextran co-polymers, agarose-polyacrylamide blends, other polymers and co-polymers with various functional groups, methacrylate derivatives and co-polymers, polystyrene and polystyrene copolymers (*see, e.g.*, Merrifield, *Biochemistry*, 3:1385-1390 (1964); Berg et al., in *Innovation Perspect. Solid Phase Synth. Collect. Pap.*, Int. Symp., 1st, Epton, Roger (Ed), pp. 453-459 (1990); Berg et al., *Pept., Proc. Eur. Pept. Symp.*, 20th, Jung, G.  
30 et al. (Eds), pp. 196-198 (1989); Berg et al., *J. Am. Chem. Soc.*, 111:8024-8026 (1989); Kent et al., *Isr. J. Chem.*, 17:243-247 (1979); Kent et

-117-

al., *J. Org. Chem.*, 43:2845-2852 (1978); Mitchell et al., *Tetrahedron Lett.*, 42:3795-3798 (1976); U.S. Patent No. 4,507,230; U.S. Patent No. 4,006,117; and U.S. Patent No. 5,389,449). Such materials include those made from polymers and co-polymers such as polyvinylalcohols, acrylates and acrylic acids  
5 such as polyethylene-co-acrylic acid, polyethylene-co-methacrylic acid, polyethylene-co-ethylacrylate, polyethylene-co-methyl acrylate, polypropylene-co-acrylic acid, polypropylene-co-methyl-acrylic acid, polypropylene-co-ethylacrylate, polypropylene-co-methyl acrylate, polyethylene-co-vinyl acetate, polypropylene-co-vinyl acetate, and those containing acid anhydride groups such as  
10 polyethylene-co-maleic anhydride and polypropylene-co-maleic anhydride. Liposomes have also been used as solid supports for affinity purifications (Powell et al. *Biotechnol. Bioeng.*, 33:173 (1989)).

Numerous methods have been developed for the immobilization of proteins and other biomolecules onto solid or liquid supports (*see, e.g.,*  
15 Mosbach, *Methods in Enzymology*, 44 (1976); Weetall, *Immobilized Enzymes, Antigens, Antibodies, and Peptides*, (1975); Kennedy et al., *Solid Phase Biochemistry, Analytical and Synthetic Aspects*, Scouten, ed., pp. 253-391 (1983); *see, generally*, Affinity Techniques. *Enzyme Purification: Part B. Methods in Enzymology*, Vol. 34, ed. W. B. Jakoby, M. Wilchek, Acad. Press,  
20 N.Y. (1974); and *Immobilized Biochemicals and Affinity Chromatography, Advances in Experimental Medicine and Biology*, vol. 42, ed. R. Dunlap, Plenum Press, N.Y. (1974)).

Among the most commonly used methods are absorption and adsorption or covalent binding to the support, either directly or via a linker, such as the  
25 numerous disulfide linkages, thioether bonds, hindered disulfide bonds, and covalent bonds between free reactive groups, such as amine and thiol groups, known to those of skill in art (*see, e.g.,* the PIERCE CATALOG, ImmunoTechnology Catalog & Handbook, 1992-1993, which describes the preparation of and use of such reagents and provides a commercial source for  
30 such reagents; Wong, *Chemistry of Protein Conjugation and Cross Linking*, CRC Press (1993); *see also* DeWitt et al., *Proc. Natl. Acad. Sci. U.S.A.*, 90:6909 (1993); Zuckermann et al., *J. Am. Chem. Soc.*, 114:10646 (1992); Kurth et al.,

-118-

*J. Am. Chem. Soc.*, 116:2661 (1994); Ellman et al., *Proc. Natl. Acad. Sci. U.S.A.*, 91:4708 (1994); Sucholeiki, *Tetrahedron Ltrs.*, 35:7307 (1994); Sun Wang, *J. Org. Chem.*, 41:3258 (1976); Padwa et al., *J. Org. Chem.*, 41:3550 (1971); and Vedejs et al., *J. Org. Chem.*, 49:575 (1984), which  
5 describe photosensitive linkers).

To effect immobilization, a composition containing the protein or other biomolecule is contacted with a support material such as alumina, carbon, an ion-exchange resin, cellulose, glass or a ceramic. Fluorocarbon polymers have been used as supports to which biomolecules have been attached by adsorption  
10 (see, U.S. Patent No. 3,843,443; Published International PCT Application WO/86 03840).

#### J. Prognosis and diagnosis

CVSP17 polypeptide proteins, domains, analogs, and derivatives thereof, and encoding nucleic acids (and sequences complementary thereto), and anti-  
15 CVSP17 polypeptide antibodies, can be used in diagnostics, particularly diagnosis of cervical cancer, colon and pancreatic cancers, and possibly other cancers, including prostate, colon, ovary, cervix breast cancers. Such molecules can be used in assays, such as immunoassays, to detect, prognose, diagnose, or monitor various conditions, diseases, and disorders affecting  
20 CVSP17 polypeptide expression, or monitor the treatment thereof. For purposes herein, the presence of CVSP17s in body fluids or tumor tissues are of particular interest.

In particular, such an immunoassay is carried out by a method including contacting a sample derived from a patient with an anti-CVSP17 polypeptide  
25 antibody under conditions such that specific binding can occur, and detecting or measuring the amount of any specific binding by the antibody. Such binding of antibody, in tissue sections, can be used to detect aberrant CVSP17 polypeptide localization or aberrant (e.g., increased, decreased or absent) levels of CVSP17 polypeptide or aberrant activity if CVSP17 or aberrant processing of CVSP17. For  
30 example, antibody to CVSP17 polypeptide can be used to assay in a patient tissue or serum sample for the presence of CVSP17 polypeptide where an aberrant level of CVSP17 polypeptide is an indication of a diseased condition.

-119-

The immunoassays which can be used include but are not limited to competitive and non-competitive assay systems using techniques such as western blots, radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoprecipitation assays, precipitin  
5 reactions, gel diffusion precipitin reactions, immunodiffusion assays, agglutination assays, complement-fixation assays, immunoradiometric assays, fluorescent immunoassays and protein A immunoassays.

CVSP17 polypeptide genes and related nucleic acid sequences and subsequences, including complementary sequences, also can be used in  
10 hybridization assays. CVSP17 polypeptide nucleic acid sequences, or subsequences thereof containing about at least 8 nucleotides, generally 14 or 16 or 30 or more, generally less than 1000 or up to 100, contiguous nucleotides can be used as hybridization probes. Hybridization assays can be used to detect, prognose, diagnose, or monitor conditions, disorders, or disease states  
15 associated with aberrant changes in CVSP17 polypeptide expression and/or activity as described herein. In particular, such a hybridization assay is carried out by a method by contacting a sample containing nucleic acid with a nucleic acid probe capable of hybridizing to CVSP17 polypeptide encoding DNA or RNA, under conditions such that hybridization can occur, and detecting or measuring  
20 any resulting hybridization.

In a specific embodiment, a method of diagnosing a disease or disorder characterized by detecting an aberrant level of a CVSP17 polypeptide in a subject is provided herein by measuring the level of the DNA, RNA, protein or activity, such as protease and/or binding activity, of a CVSP17 polypeptide in a  
25 sample derived from the subject. An increase or decrease in the level of the DNA, RNA, protein or functional activity of the CVSP17 polypeptide, relative to the level of the DNA, RNA, protein or functional activity found in an analogous sample not having the disease or disorder indicates the presence of the disease or disorder in the subject.

30 Kits for diagnostic use are also provided, that contain in one or more containers an anti-CVSP17 polypeptide antibody, and, optionally, a labeled binding partner to the antibody. Alternatively, the anti-CVSP17 polypeptide

-120-

antibody can be labeled (with a detectable marker, *e.g.*, a chemiluminescent, enzymatic, fluorescent, or radioactive moiety). A kit is also provided that includes in one or more containers a nucleic acid probe capable of hybridizing to SP protein-encoding RNA. In a specific embodiment, a kit can comprise in one  
5 or more containers a pair of primers (*e.g.*, each in the size range of 6-30 nucleotides) that are capable of priming amplification [*e.g.*, by polymerase chain reaction (see *e.g.*, Innis et al., 1990, PCR Protocols, Academic Press, Inc., San Diego, CA), ligase chain reaction (see EP 320,308) use of Q $\beta$  replicase, cyclic probe reaction, or other methods known in the art under appropriate reaction  
10 conditions of at least a portion of an SP protein-encoding nucleic acid. A kit can optionally further comprise in a container a predetermined amount of a purified CVSP17 polypeptide or nucleic acid, *e.g.*, for use as a standard or control.

**K. Pharmaceutical compositions and modes of administration**

**1. Components of the compositions**

15 Pharmaceutical compositions containing the identified compounds that modulate the activity of a CVSP17 polypeptide are provided herein. Also provided are combinations of a compound that modulates the activity of a CVSP17 polypeptide and another treatment or compound for treatment of a neoplastic disorder, such as a chemotherapeutic compound.

20 The CVSP17 polypeptide modulator and the anti-tumor agent can be packaged as separate compositions for administration together or sequentially or intermittently. Alternatively, they can provided as a single composition for administration or as two compositions for administration as a single composition. The combinations can be packaged as kits.

25 **a. CVSP17 polypeptide inhibitors**

Any CVSP17 polypeptide inhibitors, including those described herein when used alone or in combination with other compounds, that can alleviate, reduce, ameliorate, prevent, or place or maintain in a state of remission of clinical symptoms or diagnostic markers associated with neoplastic diseases,  
30 including undesired and/or uncontrolled angiogenesis, can be used in the present combinations.

-121-

For example, the CVSP17 polypeptide inhibitor is an antibody or fragment thereof that specifically reacts with a CVSP17 polypeptide or the protease domain thereof or other region thereof, such as the activation region, or is an inhibitor of the CVSP17 polypeptide production, an inhibitor of CVSP17 polypeptide membrane-localization or an inhibitor of the expression or activation of a CVSP17 polypeptide.

**b. Anti-angiogenic agents and anti-tumor agents**

Any anti-angiogenic agents and anti-tumor agents, including those described herein, when used alone or in combination with other compounds, that can alleviate, reduce, ameliorate, prevent, or place or maintain in a state of remission of clinical symptoms or diagnostic markers associated with undesired and/or uncontrolled angiogenesis and/or tumor growth and metastasis, particularly solid neoplasms, vascular malformations and cardiovascular disorders, chronic inflammatory diseases and aberrant wound repairs, circulatory disorders, crest syndromes, dermatological disorders, or ocular disorders, can be used in the combinations. Also contemplated are anti-tumor agents for use in combination with an inhibitor of a CVSP17 polypeptide.

**c. Anti-tumor agents and anti-angiogenic agents**

The compounds identified by the methods provided herein or provided herein can be used in combination with anti-tumor agents and/or anti-angiogenesis agents.

**2. Formulations and route of administration**

The compounds herein and agents can be formulated as pharmaceutical compositions, typically for single dosage administration. The concentrations of the compounds in the formulations are effective for delivery of an amount, upon administration, that is effective for the intended treatment. Typically, the compositions are formulated for single dosage administration. To formulate a composition, the weight fraction of a compound or mixture thereof is dissolved, suspended, dispersed or otherwise mixed in a selected vehicle at an effective concentration such that the treated condition is relieved or ameliorated. Pharmaceutical carriers or vehicles suitable for administration of the compounds

-122-

provided herein include any such carriers known to those skilled in the art to be suitable for the particular mode of administration.

In addition, the compounds can be formulated as the sole pharmaceutically active ingredient in the composition or can be combined with  
5 other active ingredients. Liposomal suspensions, including tissue-targeted liposomes, can also be suitable as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art. For example, liposome formulations can be prepared as described in U.S. Patent No. 4,522,811.

10 The active compound is included in the pharmaceutically acceptable carrier in an amount sufficient to exert a therapeutically useful effect in the absence of undesirable side effects on the patient treated. The therapeutically effective concentration can be determined empirically by testing the compounds in known in vitro and in vivo systems, such as the assays provided herein.

15 The concentration of active compound in the drug composition depends on absorption, inactivation and excretion rates of the active compound, the physicochemical characteristics of the compound, the dosage schedule, and amount administered as well as other factors known to those of skill in the art.

Typically a therapeutically effective dosage is contemplated. The  
20 amounts administered can be on the order of 0.001 to 1 mg/ml, including about 0.005-0.05 mg/ml and about 0.01 mg/ml, of blood volume. Pharmaceutical dosage unit forms are prepared to provide from about 1 mg to about 1000 mg, including from about 10 to about 500 mg, and including about 25-75 mg of the essential active ingredient or a combination of essential ingredients per dosage  
25 unit form. The precise dosage can be empirically determined.

The active ingredient can be administered at once, or can be divided into a number of smaller doses to be administered at intervals of time. It is understood that the precise dosage and duration of treatment is a function of the disease being treated and can be determined empirically using known testing  
30 protocols or by extrapolation from in vivo or in vitro test data. It is to be noted that concentrations and dosage values can also vary with the severity of the condition to be alleviated. It is to be further understood that for any particular

-123-

subject, specific dosage regimens should be adjusted over time according to the individual need and the professional judgment of the person administering or supervising the administration of the compositions, and that the concentration ranges set forth herein are exemplary only and are not intended to limit the scope or use of the claimed compositions and combinations containing them.

Pharmaceutically acceptable derivatives include acids, salts, esters, hydrates, solvates and prodrug forms. The derivative is typically selected such that its pharmacokinetic properties are superior to the corresponding neutral compound.

Thus, effective concentrations or amounts of one or more of the compounds provided herein or pharmaceutically acceptable derivatives thereof are mixed with a suitable pharmaceutical carrier or vehicle for systemic, topical or local administration to form pharmaceutical compositions. Compounds are included in an amount effective for ameliorating or treating the disorder for which treatment is contemplated. The concentration of active compound in the composition depends on absorption, inactivation, excretion rates of the active compound, the dosage schedule, amount administered, particular formulation as well as other factors known to those of skill in the art.

Solutions or suspensions used for parenteral, intradermal, subcutaneous, or topical application can include any of the following components: a sterile diluent, such as water for injection, saline solution, fixed oil, polyethylene glycol, glycerine, propylene glycol or other synthetic solvent; antimicrobial agents, such as benzyl alcohol and methyl parabens; antioxidants, such as ascorbic acid and sodium bisulfite; chelating agents, such as ethylenediaminetetraacetic acid (EDTA); buffers, such as acetates, citrates and phosphates; and agents for the adjustment of tonicity such as sodium chloride or dextrose. Parenteral preparations can be enclosed in ampules, disposable syringes or single or multiple dose vials made of glass, plastic or other suitable material.

-124-

In instances in which the compounds exhibit insufficient solubility, methods for solubilizing compounds can be used. Such methods are known to those of skill in this art, and include, but are not limited to, using cosolvents, such as dimethylsulfoxide (DMSO), using surfactants, such as Tween®, or  
5 dissolution in aqueous sodium bicarbonate. Derivatives of the compounds, such as prodrugs of the compounds can also be used in formulating effective pharmaceutical compositions. For ophthalmic indications, the compositions are formulated in an ophthalmically acceptable carrier. For the ophthalmic uses herein, local administration, either by topical administration or by injection are  
10 contemplated. Time release formulations are also desirable. Typically, the compositions are formulated for single dosage administration, so that a single dose administers an effective amount.

Upon mixing or addition of the compound with the vehicle, the resulting mixture can be a solution, suspension, emulsion or other composition. The form  
15 of the resulting mixture depends upon a number of factors, including the intended mode of administration and the solubility of the compound in the selected carrier or vehicle. If necessary, pharmaceutically acceptable salts or other derivatives of the compounds are prepared.

The compound is included in the pharmaceutically acceptable carrier in an  
20 amount sufficient to exert a therapeutically useful effect in the absence of undesirable side effects on the patient treated. It is understood that number and degree of side effects depends upon the condition for which the compounds are administered. For example, certain toxic and undesirable side effects are tolerated when treating life-threatening illnesses that would not be tolerated  
25 when treating disorders of lesser consequence.

The compounds also can be mixed with other active materials, that do not impair the desired action, or with materials that supplement the desired action known to those of skill in the art. The formulations of the compounds and agents for use herein include those suitable for oral, rectal, topical,  
30 inhalational, buccal (*e.g.*, sublingual), parenteral (*e.g.*, subcutaneous, intramuscular, intradermal, or intravenous), transdermal administration or any route. The most suitable route in any given case depends on the nature and

-125-

severity of the condition being treated and on the nature of the particular active compound which is being used. The formulations are provided for administration to humans and animals in unit dosage forms, such as tablets, capsules, pills, powders, granules, sterile parenteral solutions or suspensions, and oral solutions or suspensions, and oil-water emulsions containing suitable quantities of the compounds or pharmaceutically acceptable derivatives thereof. The pharmaceutically therapeutically active compounds and derivatives thereof are typically formulated and administered in unit-dosage forms or multiple-dosage forms. Unit-dose forms as used herein refers to physically discrete units suitable for human and animal subjects and packaged individually as is known in the art. Each unit-dose contains a predetermined quantity of the therapeutically active compound sufficient to produce the desired therapeutic effect, in association with the required pharmaceutical carrier, vehicle or diluent. Examples of unit-dose forms include ampoules and syringes and individually packaged tablets or capsules. Unit-dose forms can be administered in fractions or multiples thereof. A multiple-dose form is a plurality of identical unit-dosage forms packaged in a single container to be administered in segregated unit-dose form. Examples of multiple-dose forms include vials, bottles of tablets or capsules or bottles of pints or gallons. Hence, multiple dose form is a multiple of unit-doses which are not segregated in packaging.

The composition can contain along with the active ingredient: a diluent such as lactose, sucrose, dicalcium phosphate, or carboxymethylcellulose; a lubricant, such as magnesium stearate, calcium stearate and talc; and a binder such as starch, natural gums, such as gum acacia, gelatin, glucose, molasses, polvinylpyrrolidone, celluloses and derivatives thereof, povidone, crospovidones and other such binders known to those of skill in the art. Liquid pharmaceutically administrable compositions can, for example, be prepared by dissolving, dispersing, or otherwise mixing an active compound as defined above and optional pharmaceutical adjuvants in a carrier, such as, for example, water, saline, aqueous dextrose, glycerol, glycols, ethanol, and the like, to thereby form a solution or suspension. If desired, the pharmaceutical composition to be administered can also contain minor amounts of nontoxic auxiliary substances

-126-

such as wetting agents, emulsifying agents, or solubilizing agents, pH buffering agents and the like, for example, acetate, sodium citrate, cyclodextrine derivatives, sorbitan monolaurate, triethanolamine sodium acetate, triethanolamine oleate, and other such agents. Methods of preparing such dosage forms are known, or will be apparent, to those skilled in this art (see, *e.g.*, Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., 15th Edition, 1975). The composition or formulation to be administered contains a quantity of the active compound in an amount sufficient to alleviate the symptoms of the treated subject.

10 Dosage forms or compositions containing active ingredient in the range of 0.005% to 100% with the balance made up from non-toxic carrier can be prepared. For oral administration, the pharmaceutical compositions can take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*,  
15 pregelatinized maize starch, polyvinyl pyrrolidone or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets can be coated by methods well-  
20 known in the art.

The pharmaceutical preparation can also be in liquid form, for example, solutions, syrups or suspensions, or can be presented as a drug product for reconstitution with water or other suitable vehicle before use. Such liquid preparations can be prepared by conventional means with pharmaceutically  
25 acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid).

30 Formulations suitable for rectal administration can be presented as unit dose suppositories. These can be prepared by admixing the active compound

-127-

with one or more conventional solid carriers, for example, cocoa butter, and then shaping the resulting mixture.

Formulations suitable for topical application to the skin or to the eye generally are formulated as an ointment, cream, lotion, paste, gel, spray, aerosol and oil. Carriers which can be used include vaseline, lanoline, polyethylene glycols, alcohols, and combinations of two or more thereof. The topical formulations can further advantageously contain 0.05 to 15 percent by weight of thickeners selected from among hydroxypropyl methyl cellulose, methyl cellulose, polyvinylpyrrolidone, polyvinyl alcohol, poly (alkylene glycols), poly/hydroxyalkyl, (meth)acrylates or poly(meth)acrylamides. A topical formulation is often applied by instillation or as an ointment into the conjunctival sac. It also can be used for irrigation or lubrication of the eye, facial sinuses, and external auditory meatus. It can also be injected into the anterior eye chamber and other places. The topical formulations in the liquid state can be also present in a hydrophilic three-dimensional polymer matrix in the form of a strip, contact lens, and the like from which the active components are released.

For administration by inhalation, the compounds for use herein can be delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol, the dosage unit can be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin, for use in an inhaler or insufflator can be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

Formulations suitable for buccal (sublingual) administration include, for example, lozenges containing the active compound in a flavored base, usually sucrose and acacia or tragacanth; and pastilles containing the compound in an inert base such as gelatin and glycerin or sucrose and acacia.

The compounds can be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection can be presented in unit dosage form, *e.g.*, in ampules or in multi-dose

-128-

containers, with an added preservative. The compositions can be suspensions, solutions or emulsions in oily or aqueous vehicles, and can contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient can be in powder form for reconstitution with a suitable  
5 vehicle, *e.g.*, sterile pyrogen-free water or other solvents, before use.

Formulations suitable for transdermal administration can be presented as discrete patches adapted to remain in intimate contact with the epidermis of the recipient for a prolonged period of time. Such patches suitably contain the active compound as an optionally buffered aqueous solution of, for example, 0.1  
10 to 0.2 M concentration with respect to the active compound. Formulations suitable for transdermal administration can also be delivered by iontophoresis (*see, e.g.*, *Pharmaceutical Research* 3 (6), 318 (1986)) and typically take the form of an optionally buffered aqueous solution of the active compound.

The pharmaceutical compositions can also be administered by controlled  
15 release means and/or delivery devices (*see, e.g.*, in U.S. Patent Nos. 3,536,809; 3,598,123; 3,630,200; 3,845,770; 3,847,770; 3,916,899; 4,008,719; 4,687,610; 4,769,027; 5,059,595; 5,073,543; 5,120,548; 5,354,566; 5,591,767; 5,639,476; 5,674,533 and 5,733,566).

Desirable blood levels can be maintained by a continuous infusion of the  
20 active agent as ascertained by plasma levels. It should be noted that the attending physician would know how to and when to terminate, interrupt or adjust therapy to lower dosage due to toxicity, or bone marrow, liver or kidney dysfunctions. Conversely, the attending physician would also know how to and when to adjust treatment to higher levels if the clinical response is not adequate  
25 (precluding toxic side effects).

The efficacy and/or toxicity of the CVSP17 polypeptide inhibitor(s), alone or in combination with other agents also can be assessed by the methods known in the art (See generally, O'Reilly, *Investigational New Drugs*, 15:5-13 (1997)).

The active compounds or pharmaceutically acceptable derivatives can be  
30 prepared with carriers that protect the compound against rapid elimination from the body, such as time release formulations or coatings.

-129-

Kits containing the compositions and/or the combinations with instructions for administration thereof are provided. The kit can further include a needle or syringe, typically packaged in sterile form, for injecting the complex, and/or a packaged alcohol pad. Instructions are optionally included for  
5 administration of the active agent by a clinician or by the patient.

Finally, the compounds or CVSP17 polypeptides or protease domains thereof or compositions containing any of the preceding agents can be packaged as articles of manufacture containing packaging material, a compound or suitable derivative thereof provided herein, which is effective for treatment of diseases or  
10 disorders contemplated herein, within the packaging material, and a label that indicates that the compound or a suitable derivative thereof is for treating the diseases or disorders contemplated herein. The label can optionally include the disorders for which the therapy is warranted.

**L. Methods of treatment**

15 The compounds identified by the methods herein are used for treating or preventing neoplastic diseases in an animal, particularly a mammal, including a human, and are provided herein. In one embodiment, the method includes administering to a mammal an effective amount of an inhibitor of a CVSP17 polypeptide, whereby the disease or disorder is treated or prevented.

20 In an embodiment, the CVSP17 polypeptide inhibitor used in the treatment or prevention is administered with a pharmaceutically acceptable carrier or excipient. The mammal treated can be a human. The inhibitors provided herein are those identified by the screening assays. In addition, antibodies and antisense nucleic acids or double-stranded RNA (dsRNA), such as  
25 RNAi, are contemplated.

The treatment or prevention method can further include administering an anti-angiogenic treatment or agent or anti-tumor agent simultaneously with, prior to or subsequent to the CVSP17 polypeptide inhibitor, which can be any compound identified that inhibits the activity of an CVSP17 polypeptide. Such  
30 compounds include small molecule modulators, a natural product or derivative thereof, an antibody or a fragment or derivative thereof containing a binding region thereof against the CVSP17 polypeptide, an antisense nucleic acid or

-130-

double-stranded RNA (dsRNA), such as RNAi, encoding a portion of the CVSP17 polypeptide (or the complement thereof), and a nucleic acid containing at least a portion of a gene encoding the CVSP17 polypeptide into which a heterologous nucleotide sequence has been inserted such that the heterologous sequence

5 inactivates the biological activity of at least a portion of the gene encoding the CVSP17 polypeptide, in which the portion of the gene encoding a CVSP17 polypeptide flanks the heterologous sequence to promote homologous recombination with a genomic gene (or endogenous gene) encoding a CVSP17 polypeptide. In addition, such molecules are generally less than about 1000 nt

10 long.

#### 1. Antisense treatment

In a specific embodiment, as described hereinabove, CVSP17 polypeptide function is reduced or inhibited by CVSP17 polypeptide antisense nucleic acids, to treat or prevent neoplastic disease. The therapeutic or prophylactic use of

15 nucleic acids of at least six nucleotides that are antisense to a gene or cDNA encoding CVSP17 polypeptide or a portion thereof. A CVSP17 polypeptide "antisense" nucleic acid as used herein refers to a nucleic acid capable of hybridizing to a portion of a CVSP17 polypeptide RNA (generally mRNA) by virtue of some sequence complementarity, and generally under high stringency

20 conditions. The antisense nucleic acid can be complementary to a coding and/or noncoding region of a CVSP17 polypeptide mRNA. Such antisense nucleic acids have utility as therapeutics that reduce or inhibit CVSP17 polypeptide function, and can be used in the treatment or prevention of disorders as described *supra*.

The CVSP17 polypeptide antisense nucleic acids are of at least six

25 nucleotides and are generally oligonucleotides (ranging from 6 to about 150 nucleotides including 6 to 50 nucleotides). The antisense molecule can be complementary to all or a portion of the protease domain. For example, the oligonucleotide is at least 10 nucleotides, at least 15 nucleotides, at least 100 nucleotides, or at least 125 nucleotides. The oligonucleotides can be DNA or

30 RNA or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety and/or phosphate backbone. The oligonucleotide can

-131-

include other appending groups such as peptides, or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger et al., *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556 (1989); Lemaitre et al., *Proc. Natl. Acad. Sci. U.S.A.* 84:648-652 (1987); PCT Publication No. WO 88/09810, published December 5 15, 1988) or blood-brain barrier (see, *e.g.*, PCT Publication No. WO 89/10134, published April 25, 1988), hybridization-triggered cleavage agents (see, *e.g.*, Krol et al., *BioTechniques* 6:958-976 (1988)) or intercalating agents (see, *e.g.*, Zon, *Pharm. Res.* 5:539-549 (1988)).

10 The CVSP17 polypeptide antisense nucleic acid generally is an oligo-nucleotide, typically single-stranded DNA or RNA or an analog thereof or mixtures thereof. For example, the oligonucleotide includes a sequence antisense to a portion of human CVSP17 polypeptide. The oligonucleotide can be modified at any position on its structure with substituents generally known in the art.

15 The CVSP17 polypeptide antisense oligonucleotide can include at least one modified base moiety which is selected from the group including, but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, 20 dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 25 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

30 In another embodiment, the oligonucleotide includes at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose. The oligonucleotide can include at least

-132-

one modified phosphate backbone selected from a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

- 5           The oligonucleotide can be an  $\alpha$ -anomeric oligonucleotide. An  $\alpha$ -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which the strands run parallel to each other (Gautier et al., *Nucl. Acids Res.* 15:6625-6641 (1987)).

- The oligonucleotide can be conjugated to another molecule, *e.g.*, a  
10   peptide, hybridization triggered cross-linking agent, transport agent and hybridization-triggered cleavage agent.

- The oligonucleotides can be synthesized by standard methods known in the art, *e.g.* by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples,  
15   phosphorothioate oligonucleotides can be synthesized by the method of Stein et al. (*Nucl. Acids Res.* 16:3209 (1988)), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., *Proc. Natl. Acad. Sci. U.S.A.* 85:7448-7451 (1988)), etc.

- In a specific embodiment, the CVSP17 polypeptide antisense  
20   oligonucleotide includes catalytic RNA or a ribozyme (see, *e.g.*, PCT International Publication WO 90/11364, published October 4, 1990; Sarver et al., *Science* 247:1222-1225 (1990)). In another embodiment, the oligonucleotide is a 2'-O-methylribonucleotide (Inoue et al., *Nucl. Acids Res.* 15:6131-6148 (1987)), or a chimeric RNA-DNA analogue (Inoue et al., *FEBS Lett.* 215:327-330 (1987)).  
25   Alternatively, the oligonucleotide can be double-stranded RNA (dsRNA) such as RNAi.

- In an alternative embodiment, the CVSP17 polypeptide antisense nucleic acid is produced intracellularly by transcription from an exogenous sequence. For example, a vector can be introduced *in vivo* such that it is taken up by a cell,  
30   within which cell the vector or a portion thereof is transcribed, producing an antisense nucleic acid (RNA). Such a vector would contain a sequence encoding the CVSP17 polypeptide antisense nucleic acid. Such a vector can remain

-133-

episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the CVSP17 polypeptide antisense RNA can be by any promoter known in the art to act in mammalian, including human, cells. Such promoters can be inducible or constitutive. Such promoters include but are not limited to: the SV40 early promoter region (Bernoist and Chambon, *Nature* 290:304-310 (1981), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto et al., *Cell* 22:787-797 (1980), the herpes thymidine kinase promoter (Wagner et al., *Proc. Natl. Acad. Sci. U.S.A.* 78:1441-1445 (1981), the regulatory sequences of the metallothionein gene (Brinster et al., *Nature* 296:39-42 (1982), etc.

The antisense nucleic acids include sequence complementary to at least a portion of an RNA transcript of a CVSP17 polypeptide gene, including a human CVSP17 polypeptide gene. Absolute complementarity is not required.

The amount of CVSP17 polypeptide antisense nucleic acid that is effective in the treatment or prevention of neoplastic disease depends on the nature of the disease, and can be determined empirically by standard clinical techniques. Where possible, it is desirable to determine the antisense cytotoxicity in cells *in vitro*, and then in useful animal model systems prior to testing and use in humans.

## 2. RNA interference

RNA interference (RNAi) (see, *e.g.* Chuang et al. (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:4985) can be employed to inhibit the expression of a gene encoding an CVSP17. Interfering RNA (RNAi) fragments, particularly double-stranded (ds) RNAi, can be used to generate loss-of-CVSP17 function. Methods relating to the use of RNAi to silence genes in organisms including, mammals, *C. elegans*, *Drosophila* and plants, and humans are known (see, *e.g.*, Fire et al. (1998) *Nature* 391:806-811; Fire (1999) *Trends Genet.* 15:358-363; Sharp (2001) *Genes Dev.* 15:485-490; Hammond, et al. (2001) *Nature Rev. Genet.* 2:110-1119; Tuschl (2001) *Chem. Biochem.* 2:239-245; Hamilton et al.

-134-

(1999) *Science* 286:950-952; Hammond *et al.* (2000) *Nature* 404:293-296; Zamore *et al.* (2000) *Cell* 101:25-33; Bernstein *et al.* (2001) *Nature* 409: 363-366; Elbashir *et al.* (2001) *Genes Dev.* 15:188-200; Elbashir *et al.* (2001) *Nature* 411:494-498; International PCT application No. WO 01/29058; International  
5 PCT application No. WO 99/32619). By selecting appropriate sequences, expression of dsRNA can interfere with accumulation of endogenous mRNA encoding an CVSP17.

Double-stranded RNA (dsRNA)-expressing constructs are introduced into a host, such as an animal or plant. This can be accomplished by any of  
10 numerous methods known in the art, for example by including it in a replicable vector, such as a viral vector (see discussion below), that remains episomal or integrates into the genome. The dsRNA can be introduced into an appropriate nucleic acid expression vector and administering it so that it becomes intracellular, *e.g.*, by infection using a defective or attenuated retroviral or other  
15 viral vector (see U.S. Patent No. 4,980,286). Other methods include, but are not limited to, direct injection of naked DNA, using microparticle bombardment (*e.g.*, a gene gun; Biolistic, Dupont), coating with lipids or cell-surface receptors or transfecting agents, encapsulation in liposomes, microparticles, or microcapsules, administering it in linkage to a peptide which is known to enter  
20 the nucleus, administering it in linkage to a ligand subject to receptor-mediated endocytosis (see *e.g.*, Wu and Wu, *J. Biol. Chem.* 262:4429-4432 (1987)) (which can be used to target cell types specifically expressing the receptors) and other methods. In other methods, a nucleic acid-ligand complex can be formed in which the ligand is a fusogenic viral peptide to disrupt endosomes, allowing  
25 the nucleic acid to avoid lysosomal degradation. In other methods, the nucleic acid can be targeted *in vivo* for cell specific uptake and expression, by targeting a specific receptor (see, *e.g.*, PCT Publications WO 92/06180 dated April 16, 1992 (Wu *et al.*); WO 92/22635 dated December 23, 1992 (Wilson *et al.*); WO92/20316 dated November 26, 1992 (Findeis *et al.*); WO93/14188 dated  
30 July 22, 1993 (Clarke *et al.*), WO 93/20221 dated October 14, 1993 (Young)). Alternatively, the nucleic acid can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination (Koller and

-135-

Smithies, *Proc. Natl. Acad. Sci. USA* 86:8932-8935 (1989); Zijlstra et al., *Nature* 342:435-438 (1989)).

RNAi can be used to inhibit expression *in vitro* or *in vivo*. Regions include at least about 21 (or 21) nucleotides that are selective (i.e. unique) for CVSP17  
5 are used to prepare the RNAi. Smaller fragments of about 21 nucleotides can be transformed directly (i.e., *in vitro* or *in vivo*) into cells; larger RNAi dsRNA molecules are generally introduced using vectors that encode them. dsRNA molecules are at least about 21 bp long or longer, such as 50, 100, 150, 200 and longer. Methods, reagents and protocols for introducing nucleic acid  
10 molecules in to cells *in vitro* and *in vivo* are known to those of skill in the art.

### 3. Gene Therapy

In an exemplary embodiment, nucleic acids that include a sequence of nucleotides encoding a CVSP17 polypeptide or functional domains or derivative thereof, are administered to promote CVSP17 polypeptide function, by way of  
15 gene therapy. In this embodiment, the nucleic acid produces an encoded protein (or the nucleic acid or encoded RNA) that mediates a therapeutic effect by promoting CVSP17 polypeptide function. Any of the methods for gene therapy available in the art can be used (see, Goldspiel et al., *Clinical Pharmacy* 12:488-505 (1993); Wu and Wu, *Biotherapy* 3:87-95 (1991); Tolstoshev, *An. Rev. Pharmacol. Toxicol.* 32:573-596 (1993); Mulligan, *Science* 260:926-932 (1993); and Morgan and Anderson, *An. Rev. Biochem.* 62:191-217 (1993); *TIBTECH* 11(5):155-215 (1993). For example, one therapeutic composition for  
20 gene therapy includes a CVSP17 polypeptide-encoding nucleic acid that is part of an expression vector that expresses a CVSP17 polypeptide or domain, fragment or chimeric protein thereof in a suitable host. In particular, such a nucleic acid has a promoter operably linked to the CVSP17 polypeptide coding region, the promoter being inducible or constitutive, and, optionally, tissue-specific. In another particular embodiment, a nucleic acid molecule is used in which the CVSP17 polypeptide coding sequences and any other desired  
25 sequences are flanked by regions that promote homologous recombination at a desired site in the genome, thus providing for intrachromosomal expression of  
30

-136-

the SP protein nucleic acid (Koller and Smithies, *Proc. Natl. Acad. Sci. USA* 86:8932-8935 (1989); Zijlstra et al., *Nature* 342:435-438 (1989)).

Delivery of the nucleic acid into a patient can be either direct, in which case the patient is directly exposed to the nucleic acid or nucleic acid-carrying  
5 vector, or indirect, in which case, cells are first transformed with the nucleic acid *in vitro*, then transplanted into the patient. These two approaches are known, respectively, as *in vivo* or *ex vivo* gene therapy.

In a specific embodiment, the nucleic acid is directly administered *in vivo*, and it is expressed to produce the encoded product. This can be accomplished  
10 by any of numerous methods known in the art, *e.g.*, by constructing it as part of an appropriate nucleic acid expression vector and administering it so that it becomes intracellular, *e.g.*, by infection using a defective or attenuated retroviral or other viral vector (see U.S. Patent No. 4,980,286), or by direct injection of naked DNA, or by use of microparticle bombardment (*e.g.*, a gene gun; Biolistic,  
15 Dupont), or coating with lipids or cell-surface receptors or transfecting agents, encapsulation in liposomes, microparticles, or microcapsules, or by administering it in linkage to a peptide which is known to enter the nucleus, by administering it in linkage to a ligand subject to receptor-mediated endocytosis (see *e.g.*, Wu and Wu, *J. Biol. Chem.* 262:4429-4432 (1987)) (which can be used to target cell  
20 types specifically expressing the receptors), etc. In another embodiment, a nucleic acid-ligand complex can be formed in which the ligand is a fusogenic viral peptide to disrupt endosomes, allowing the nucleic acid to avoid lysosomal degradation. In yet another embodiment, the nucleic acid can be targeted *in vivo* for cell specific uptake and expression, by targeting a specific receptor (see,  
25 *e.g.*, PCT Publications WO 92/06180 dated April 16, 1992 (Wu et al.); WO 92/22635 dated December 23, 1992 (Wilson et al.); WO92/20316 dated November 26, 1992 (Findeis et al.); WO93/14188 dated July 22, 1993 (Clarke et al.), WO 93/20221 dated October 14, 1993 (Young)). Alternatively, the nucleic acid can be introduced intracellularly and incorporated within host cell  
30 DNA for expression, by homologous recombination (Koller and Smithies, *Proc. Natl. Acad. Sci. USA* 86:8932-8935 (1989); Zijlstra et al., *Nature* 342:435-438 (1989)).

-137-

In a specific embodiment, a viral vector that contains the CVSP17 polypeptide nucleic acid is used. For example, a retroviral vector can be used (see Miller et al., *Meth. Enzymol.* 217:581-599 (1993)). These retroviral vectors have been modified to delete retroviral sequences that are not necessary for packaging of the viral genome and integration into host cell DNA. The CVSP17 polypeptide nucleic acid to be used in gene therapy is cloned into the vector, which facilitates delivery of the gene into a patient. More detail about retroviral vectors can be found in Boesen et al., *Biotherapy* 6:291-302 (1994), which describes the use of a retroviral vector to deliver the *mdr1* gene to hematopoietic stem cells in order to make the stem cells more resistant to chemotherapy. Other references illustrating the use of retroviral vectors in gene therapy are: Clowes et al., *J. Clin. Invest.* 93:644-651 (1994); Kiem et al., *Blood* 83:1467-1473 (1994); Salmons and Gunzberg, *Human Gene Therapy* 4:129-141 (1993); and Grossman and Wilson, *Curr. Opin. in Genetics and Devel.* 3:110-114 (1993).

Adenoviruses are other viral vectors that can be used in gene therapy. Adenoviruses are especially attractive vehicles for delivering genes to respiratory epithelia. Adenoviruses naturally infect respiratory epithelia where they cause a mild disease. Other targets for adenovirus-based delivery systems are liver, the central nervous system, endothelial cells, and muscle. Adenoviruses have the advantage of being capable of infecting non-dividing cells. Kozarsky and Wilson, *Current Opinion in Genetics and Development* 3:499-503 (1993) present a review of adenovirus-based gene therapy. Bout et al., *Human Gene Therapy* 5:3-10 (1994) demonstrated the use of adenovirus vectors to transfer genes to the respiratory epithelia of rhesus monkeys. Other instances of the use of adenoviruses in gene therapy can be found in Rosenfeld et al., *Science* 252:431-434 (1991); Rosenfeld et al., *Cell* 68:143-155 (1992); and Mastrangeli et al., *J. Clin. Invest.* 91:225-234 (1993). Adeno-associated virus (AAV) also is used in gene therapy (Walsh et al., *Proc. Soc. Exp. Biol. Med.* 204:289-300 (1993)).

Another approach to gene therapy involves transferring a gene to cells in tissue culture by such methods as electroporation, lipofection, calcium phosphate mediated transfection, or viral infection. Usually, the method of

-138-

transfer includes the transfer of a selectable marker to the cells. The cells are then placed under selection to isolate those cells that have taken up and are expressing the transferred gene. Those cells are then delivered to a patient.

In this embodiment, the nucleic acid is introduced into a cell prior to  
5 administration *in vivo* of the resulting recombinant cell. Such introduction can be carried out by any method known in the art, including but not limited to transfection, electroporation, microinjection, infection with a viral or bacteriophage vector containing the nucleic acid sequences, cell fusion, chromosome-mediated gene transfer, microcell-mediated gene transfer,  
10 spheroplast fusion, etc. Numerous techniques are known in the art for the introduction of foreign genes into cells (see *e.g.*, Loeffler and Behr, *Meth. Enzymol.* 217:599-618 (1993); Cohen et al., *Meth. Enzymol.* 217:618-644 (1993); Cline, *Pharmac. Ther.* 29:69-92 (1985)) and can be used, provided that the necessary developmental and physiological functions of the recipient cells  
15 are not disrupted. The technique should provide for the stable transfer of the nucleic acid to the cell, so that the nucleic acid is expressible by the cell and generally heritable and expressible by its cell progeny.

The resulting recombinant cells can be delivered to a patient by various methods known in the art. In an embodiment, epithelial cells are injected, *e.g.*,  
20 subcutaneously. In another embodiment, recombinant skin cells can be applied as a skin graft onto the patient. Recombinant blood cells (*e.g.*, hematopoietic stem or progenitor cells) can be administered intravenously. The amount of cells envisioned for use depends on the desired effect, patient state, etc., and can be determined by one skilled in the art.

25 Cells into which a nucleic acid can be introduced for purposes of gene therapy encompass any desired, available cell type, and include but are not limited to epithelial cells, endothelial cells, keratinocytes, fibroblasts, muscle cells, hepatocytes; blood cells such as T lymphocytes, B lymphocytes, monocytes, macrophages, neutrophils, eosinophils, megakaryocytes,  
30 granulocytes; various stem or progenitor cells, in particular hematopoietic stem or progenitor cells, *e.g.*, such as stem cells obtained from bone marrow, umbilical cord blood, peripheral blood, fetal liver, and other sources thereof.

-139-

For example, a cell used for gene therapy is autologous to the patient. In an embodiment in which recombinant cells are used in gene therapy, a CVSP17 polypeptide nucleic acid is introduced into the cells such that it is expressible by the cells or their progeny, and the recombinant cells are then administered *in vivo* for therapeutic effect. In a specific embodiment, stem or progenitor cells are used. Any stem and/or progenitor cells which can be isolated and maintained *in vitro* can potentially be used in accordance with this embodiment. Such stem cells include but are not limited to hematopoietic stem cells (HSC), stem cells of epithelial tissues such as the skin and the lining of the gut, embryonic heart muscle cells, liver stem cells (PCT Publication WO 94/08598, dated April 28, 1994), and neural stem cells (Stemple and Anderson, *Cell* 71:973-985 (1992)).

Epithelial stem cells (ESCs) or keratinocytes can be obtained from tissues such as the skin and the lining of the gut by known procedures (Rheinwald, *Meth. Cell Bio.* 21A:229 (1980)). In stratified epithelial tissue such as the skin, renewal occurs by mitosis of stem cells within the germinal layer, the layer closest to the basal lamina. Stem cells within the lining of the gut provide for a rapid renewal rate of this tissue. ESCs or keratinocytes obtained from the skin or lining of the gut of a patient or donor can be grown in tissue culture (Rheinwald, *Meth. Cell Bio.* 21A:229 (1980); Pittelkow and Scott, *Cano Clinic Proc.* 61:771 (1986)). If the ESCs are provided by a donor, a method for suppression of host versus graft reactivity (*e.g.*, irradiation, drug or antibody administration to promote moderate immunosuppression) also can be used.

With respect to hematopoietic stem cells (HSC), any technique which provides for the isolation, propagation, and maintenance *in vitro* of HSC can be used in this embodiment. Techniques by which this can be accomplished include (a) the isolation and establishment of HSC cultures from bone marrow cells isolated from the future host, or a donor, or (b) the use of previously established long-term HSC cultures, which can be allogeneic or xenogeneic. Non-autologous HSC generally are used with a method of suppressing transplantation immune reactions of the future host/patient. In a particular embodiment, human bone marrow cells can be obtained from the posterior iliac

-140-

crest by needle aspiration (see, *e.g.*, Kodo et al., *J. Clin. Invest.* 73:1377-1384 (1984)). For example, the HSCs can be made highly enriched or in substantially pure form. This enrichment can be accomplished before, during, or after long-term culturing, and can be done by any techniques known in the art. Long-term  
5 cultures of bone marrow cells can be established and maintained by using, for example, modified Dexter cell culture techniques (Dexter et al., *J. Cell Physiol.* 91:335 (1977) or Witlock-Witte culture techniques (Witlock and Witte, *Proc. Natl. Acad. Sci. USA* 79:3608-3612 (1982)).

In a specific embodiment, the nucleic acid to be introduced for purposes  
10 of gene therapy includes an inducible promoter operably linked to the coding region, such that expression of the nucleic acid is controllable by controlling the presence or absence of the appropriate inducer of transcription.

### 3. Prodrugs

A method for treating tumors is provided. The method is practiced by  
15 administering a prodrug that is cleaved at a specific site by an CVSP17 to release an active drug or a precursor that can be converted to active drug *in vivo*. Upon contact with a cell that expresses CVSP17 activity, the prodrug is converted into an active drug. The prodrug can be a conjugate that contains the active agent, such as an anti-tumor drug, such as a cytotoxic agent, or other  
20 therapeutic agent (TA), linked to a substrate for the targeted CVSP17, such that the drug or agent is inactive or unable to enter a cell, in the conjugate, but is activated upon cleavage. The prodrug, for example, can contain an oligopeptide, typically a relatively short, less than about 10 amino acids peptide, that is proteolytically cleaved by the targeted CVSP17. Cytotoxic agents, include, but  
25 are not limited to, alkylating agents, antiproliferative agents and tubulin binding agents. Others include, vinca drugs, mitomycins, bleomycins and taxanes.

### M. Animal models

Transgenic animal models and animals, such as rodents, including mice and rats, cows, chickens, pigs, goats, sheep, monkeys, including gorillas, and  
30 other primates, are provided herein. In particular, transgenic non-human animals that contain heterologous nucleic acid encoding an CVSP17 polypeptide or a transgenic animal in which expression of the polypeptide has been altered, such

-141-

as by replacing or modifying the promoter region or other regulatory region of the endogenous gene are provided. Such an animal can be produced by promoting recombination between endogenous nucleic acid and an exogenous CVSP17 gene that could be over-expressed or mis-expressed, such as by  
5 expression under a strong promoter, via homologous or other recombination event.

Transgenic animals can be produced by introducing the nucleic acid using any known method of delivery, including, but not limited to, microinjection, lipofection and other modes of gene delivery into a germline cell or somatic cells,  
10 such as an embryonic stem cell. Typically the nucleic acid is introduced into a cell, such as an embryonic stem cell (ES), followed by injecting the ES cells into a blastocyst, and implanting the blastocyst into a foster mother, which is followed by the birth of a transgenic animal. Generally, introduction of a heterologous nucleic acid molecule into a chromosome of the animal occurs by a  
15 recombination between the heterologous CVSP17-encoding nucleic acid and endogenous nucleic acid. The heterologous nucleic acid can be targeted to a specific chromosome.

In some instances, knockout animals can be produced. Such an animal can be initially produced by promoting homologous recombination between an  
20 CVSP17 polypeptide gene in its chromosome and an exogenous CVSP17 polypeptide gene that has been rendered biologically inactive (typically by insertion of a heterologous sequence, *e.g.*, an antibiotic resistance gene). In one embodiment, this homologous recombination is performed by transforming embryo-derived stem (ES) cells with a vector containing the insertionally  
25 inactivated CVSP17 polypeptide gene, such that homologous recombination occurs, followed by injecting the ES cells into a blastocyst, and implanting the blastocyst into a foster mother, followed by the birth of the chimeric animal ("knockout animal") in which an CVSP17 polypeptide gene has been inactivated (see Capecchi, *Science* 244:1288-1292 (1989)). The chimeric animal can be  
30 bred to produce homozygous knockout animals, which can then be used to produce additional knockout animals. Knockout animals include, but are not limited to, mice, hamsters, sheep, pigs, cattle, and other non-human mammals.

-142-

For example, a knockout mouse is produced. The resulting animals can serve as models of specific diseases, such as cancers, that exhibit under-expression of an CVSP17 polypeptide. Such knockout animals can be used as animal models of such diseases *e.g.*, to screen for or test molesules for the ability to treat or  
5 prevent such diseases or disorders.

Other types of transgenic animals also can be produced, including those that over-express the CVSP17 polypeptide. Such animals include "knock-in" animals that are animals in which the normal gene is replaced by a variant, such as a mutant, an over-expressed form, or other form. For example, one species',  
10 such as a rodent's endogenous gene can be replaced by the gene from another species, such as from a human. Animals also can be produced by non-homologous recombination into other sites in a chromosome; including animals that have a plurality of integration events.

After production of the first generation transgenic animal, a chimeric  
15 animal can be bred to produce additional animals with over-expressed or mis-expressed CVSP17 polypeptides. Such animals include, but are not limited to, mice, hamsters, sheep, pigs, cattle and other non-human mammals. The resulting animals can serve as models of specific diseases, such as cancers, that exhibit over-expression or mis-expression of an CVSP17 polypeptide. Such  
20 animals can be used as animal models of such diseases *e.g.*, to screen for or test molecules for the ability to treat or prevent such diseases or disorders. In a specific embodiment, a mouse with over-expressed or mis-expressed CVSP17 polypeptide is produced.

25 The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

#### EXAMPLE 1

##### Identiflcation of CVSP17

The protein sequence of the protease domain of endotheliase1 (ET1; also  
30 called DESC1, accession number XP\_003340) was used to search the human HTGS (High Throughput Genomic Sequence) database using the *tblastn* algorithm (<http://www.ncbi.nlm.nih.gov/BLAST>). This search and alignment

-143-

algorithm compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands). Several potential novel serine proteases were identified. One of them will be referred hereafter as CVSP17. CVSP17 shared 32% identity to the protease domain of ET1. A search using the algorithm *blastp* (<http://www.ncbi.nlm.nih.gov/BLAST>) indicated that the translated sequence of CVSP17 showed 41% identity to MTSP6 (see U.S. application Serial No. 09/776,191 and corresponding published International PCT application No. WO 01/57194, also later reported as TADG12 [accession number NP\_071759]) and 40% identity to human enterokinase (accession number NP\_002763.1).

Based on the incomplete and unordered human genome sequence (<http://www.ncbi.nlm.nih.gov/genome/seq>), CVSP17 appears to be localized on chromosome 2 (locus: 2q37.1; clone accession number AF307337). A search of sequences deposited in GenBank showed that no identical cDNA sequence had been deposited. A CVSP17 sequence is found within a 220-kbp genomic region of chromosome 2q37.1 sequenced by Rosenthal's group in Germany and published in *Genomics* (73: 50-55, 2001). This 220-kbp genomic region contains the genes encoding the human alkaline phosphatase, the X chromosome controlling element and the nicotinic cholinergic receptor. Interestingly Rosenthal's group did not report the presence of the CVSP17 gene encoding a novel serine protease and whose sequence is localized between the X chromosome controlling element and nicotinic cholinergic receptor genes. An earlier search of the EST database did not show the existence of any EST clone. A later search of the human EST database showed that an EST clone from human testis (IMAGE cDNA clone 5269030; GenBank accession number BI464671) includes sequence similar to the CVSP17 protease domain sequence, but this clone includes point mutations and many frameshift mutations compared to the nucleic acid molecule, and hence does not provide a CVSP17 polypeptide or a protease domain therefor.

#### 30 Cloning of CVSP17 genomic fragment from human genomic DNA

Using the electronically retrieved genomic sequence of CVSP17, two gene-specific oligonucleotide primers within an exon region of CVSP17 were

-144-

designed and synthesized (<http://www.gensetoligos.com>). The sequence for the 5' end primer is 5'-CTGAGCCTGGCCCCCGCCCTAGAGAGGTC-3' (SEQ ID No. 7) and that of the 3' end primer is 5'-GGACAGGGGTCAGCTCACCCCTCTGTTTG-3' (SEQ ID No. 8). These primers were used to amplify a 317-bp genomic fragment. The PCR product was isolated, purified using the MinElute gel extraction kit (catalog number 28606; <http://www.qiagen.com>) and subcloned into an *E. coli* vector (TOPO-TA cloning kit; catalog no. K-4500-01, <http://www.invitrogen.com>). The sequence of this genomic fragment was verified to match that of the genomic sequence of CVSP17 using a fluorescent dye-based DNA sequencing method (catalog number 4390244; ABI PRISM® BigDye™ Terminator v 3.0 Ready Reaction Cycle Sequencing Kits with AmpliTaq® DNA Polymerase, FS; <http://home.appliedbiosystems.com>).

**Gene expression profile of CVSP17 in normal, tumor tissues and cell lines**

To obtain information regarding the gene expression profile of the CVSP17 transcript, the 317-bp CVSP17 genomic fragment was used to probe a dot blot composed of polyA<sup>+</sup>RNA extracted and purified from 76 different human tissues (Human Multiple Tissue Expression (MTE) Array; catalog no. 7775-1; <http://www.clontech.com>). The results indicated that CVSP17 is strongly expressed in the cervical carcinoma cell line, HeLaS3. PCR amplification of the CVSP17 cDNA from cDNA libraries made from several human primary tumors xenografted in nude mice (human tumor multiple tissue cDNA (MTC) panel, catalog number K1522-1, <http://www.clontech.com>) was performed using CVSP17-specific primers. The CVSP17 cDNA was not detected in any of the 8 tumor samples tested, including breast carcinoma (GI-101), lung carcinomas (LX-1 & GI-117), colon adenocarcinomas (GI-112 & CX-1), pancreatic adenocarcinoma (GI-103), ovarian carcinoma (GI-102), and prostatic carcinoma (PC-3).

**Cloning of CVSP17 from HeLaS3 cell line using RACE reactions**

Using the electronically retrieved genomic sequence of CVSP17, four exonic oligonucleotide primers were designed. The sequence for the 5' end primer is 5'-GAGCCCCAGGAGCCCCCTGCCGGAACCGCC-3' (SEQ ID No. 9) and that of the 3' end primer is 5'-ACCTCTCTAGGGCGGGGGCCAGGCTCAG-3'

-145-

(SEQ ID No. 10). The sequence for the nested 5' end primer is 5'-TGGCACGAGTCAACGCCCCCGCCAGGTAC-3' (SEQ ID No. 11) and that of the nested 3' end primer is 5'-TCGCGGGCTGGGGCGCCCTCTTCGAAGACG-3' (SEQ ID No. 12). The first set of RACE primers was used to amplify cDNA fragments from human HeLaS3 Marathon-ready cDNA library (catalog number 7439-1; <http://www.clontech.com>).

Following this reaction, nested RACE reactions were performed using the nested primers. Several DNA bands were detected in all RACE reactions. To identify bands that contained CVSP17 cDNA, Southern analysis was performed using a ~260-bp cDNA probe amplified from the first set of 5'- and 3'-RACE primers on HeLaS3 cDNA library. A 0.9 kbp cDNA fragment was isolated from the nested 5'-RACE reaction and a 1.4 kbp cDNA fragment was isolated from the nested 3'-RACE reaction. These RACE products were isolated, purified using the MinElute gel extraction kit (<http://www.qiagen.com>) and subcloned into an *E. coli* vector (TOPO-TA cloning kit; <http://www.invitrogen.com>). Subsequent sequence analysis confirmed that the nucleotide sequence of these cDNA fragments matched that of the genomic CVSP17 exon sequences and also contained the missing cDNA sequences. The 5'-RACE product did not extend to the beginning of the cDNA as the methionine start codon was missing. An additional 5'-RACE reaction was performed using another primer: 5'-GCCAGCGTCACAGTCCACAGAAGCTCATTC-3' (SEQ ID No. 15). A ~0.5-kbp RACE product was isolated, subcloned as above and sequenced. Sequence analysis indicated that this RACE product contained the start codon.

To obtain the full-length CVSP17 cDNA, an end-to-end PCR amplification using gene-specific primers and the cDNA library made from human HeLaS3 was used. The two primers used were: 5'-CTGGTCACCATGCTGCTGGCTGTGCTGCTG-3' for the 5' end (SEQ ID No. 16) (start codon underlined) and 5'-GGGCAGCGACAGTTTGTTCATTATGCTCCCG-3' (SEQ ID No. 17) for the 3' end. The sequences for both primers were derived from the cDNA sequence of CVSP17 RACE products. The 3' primer corresponds to the sequence downstream of the stop codon. A ~2.1-kbp fragment was amplified from the human HeLaS3 cDNA library. The PCR product

-146-

was isolated, purified using the MinElute gel extraction kit (<http://www.qiagen.com>) and subcloned into an *E. coli* vector (TOPO-TA cloning kit; <http://www.invitrogen.com>). Sequence analysis was performed to confirm the nucleotide sequence.

## 5 Homology of CVSP17 to other serine proteases

Sequence and protein domain analyses of the translated CVSP17 protein show that CVSP17 contains a signal peptide sequence (aa 1 to aa 17) and a trypsin-like serine protease domain (aa 104 to aa 332) characterized by the presence of a protease activation cleavage site (...R↓VGGSAAPP...), where ↓  
 10 indicates the protease activation cleavage site at the beginning of the domain and the catalytic triad residues (histidine, aspartate and serine) in 3 highly-conserved regions of the catalytic domain.

CVSP17 contains a signal peptide sequence (aa 1 to aa 17) and a trypsin-like serine protease domain (aa 104 to aa 332) characterized by the presence of  
 15 a protease activation cleavage site (...R<sub>104</sub>↓I<sub>105</sub>VGGSAAPP..., where ↓ indicates putative protease activation cleavage site) at the beginning of the domain, and the catalytic triad residues (H<sub>145</sub>, D<sub>191</sub> and S<sub>286</sub>) in 3 highly-conserved regions of the catalytic domain. In addition CVSP17 has an additional 303-amino acid sequence (aa 333 to aa 635). Analysis of this 303-amino acid long region did  
 20 not match any known domain. Three leucine zipper patterns (aa 432-453; aa 439-460; aa 446-467) were identified.

CVSP17 has an *N*-linked glycosylation site (...N<sub>9</sub>VT...) and an unpaired cysteine (C<sub>211</sub>) in the protease domain that is predicted to pair with C<sub>88</sub> outside of the protease domain. The following cysteine pairings in the putative protease  
 25 domain can be noted: C<sub>130</sub>-C<sub>146</sub>; C<sub>225</sub>-C<sub>292</sub>; C<sub>256</sub>-C<sub>271</sub> and C<sub>282</sub>-C<sub>313</sub>. Alignment (*blastp*; <http://www.ncbi.nlm.nih.gov/BLAST>) of the protease domain (minus the 303-amino acid extension at the C terminus) sequence of CVSP17 with that of human enterokinase (accession number NP\_002763.1) and MTSP6 (also called TADG12; accession number NP\_071759) showed 40% and 41% identity in the  
 30 protease domain, respectively. CVSP17 also shares homology to several other serine proteases including DESC1 (37% identity; accession number XP\_003340); prostamin (37%; accession number BAB20376); matriptase (36%;

-147-

accession number NP\_068813) and airway trypsin-like protease (39%; accession number NP\_004253).

Also, International PCT application Nos. WO 02/000860, WO 02/038744, WO 02/024886, WO 01/22920 and WO 01/24815 describe polypeptides that have homology with CVSP17 polypeptides as provided herein. For example, International PCT application No. WO 01/22920 provides a polypeptide that has homology with only amino acids 58-279 of the CVSP17 of SEQ ID No. 6, and International PCT application No. WO 01/24815 provides a polypeptide that has homology amino acids 12-202. The polypeptides of the other publications also have differences from CVSP17 as provided herein. None include the C-terminal portion that includes amino acid 397-427 of SEQ ID No. 6.

#### Sequence analysis

CVSP17 cDNA and protein sequences were analyzed using MacVector (version 6.5.3; <http://www.accelrys.com/products/macvector/index.html>). The full length cDNA encoding CVSP17 is 2,173 bp long containing a 1,908-bp open reading frame, which translates to a 635-amino acid protein. The cDNA encoding the protease domain is 684 bp long which translates to a 228-amino acid domain. The G+C content of the CVSP17 cDNA is 71%. The following are the cDNA sequence and the translated protein sequence of CVSP17 (see, also SEQ ID Nos. 5 and 6). cDNA sequence of CVSP17:

CVSP17 full length cDNA and translated protein sequence  
Sequence Range: 1 to 2173

25	10	20	30	40	50	60
	CTAGAATTCAGCGCGCTGAATTC	TAGCCAGCTCCTGGTCACCATGCTGCTGGCTGTGC				
	GATCTTAAGTCGCGCGACTTAAGATCGGGTCGAGGACCAAGTGGTACGACGACCGACACG					
					M L L A V	
	70	80	90	100	110	120
30	TGCTGCTGCTACCCCTCCCAAGCTCATGGTTTGCCACGGGCACCCACTGTACACACGCC					
	ACGACGACGATGGGGAGGGTTCGAGTACCAAACGGGTGCCCGTGGGTGACATGTGTGCGG					
	L L L L P L P S S W F A H G H P L Y T R					
	130	140	150	160	170	180
	TGCCCCCAGCACCTGCAAGTTCTGTGCGGCCAGGGGACTCAGGCGTTGCAGGCAGCCC					
35	ACGGGGGGTTCGTGGGACGTTCAAGACAGCCGGTCCCTGAGTCCGCAACGTCCTCGGG					
	L P P S T L Q V L S A Q G T Q A L Q A A					
	190	200	210	220	230	240
	AGAGGAGCGCCAGTGGGCAATAAACCGAGTGGCGATGGAGATCCAGCACAGATCGCACG					
	TCTCCTCGCGGTACCCGTTATTTGGCTCACCGCTACCTCTAGGTCGTGTCTAGCGTGC					
40	Q R S A Q W A I N R V A M E I Q H R S H					
	250	260	270	280	290	300
	AGTGCAGGATCTGGGCGCCCCAGGCCCTCAAGCTCTCCTCCAGGACCCACCTGAGCCAG					
	TCACGGCTCCTAGACCGCGGGGTCCGGAGTTCGAGAGGAGGTCCTGGGTGGACTCGGTC					
	E C R G S G R P R P Q A L L Q D P P E P					
45	310	320	330	340	350	360
	GGCCGTGCGGCGAGAGCGTCCGAGCACTGCCAATGTGACGCGGGCCACGGCCGATCG					
	CCGGCACGCCGCTCTCCGACGGCTCGTGACGGTTACACTGCGCCCGGGTGCCGGCGTAGC					

-148-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

G P C G E R R P S T A N V T R A H G R I  
370 380 390 400 410 420  
TGGGGGGCAGCGCGGGCGCGCCCGGGGCTGGCCCTGGTGGTGGAGGCTGCAGCTCGGCG  
ACCCCCCGTCGCGCGCGCGGGGCCCCGGACCGGGACCGACCACTCCGACGTGAGCCGC  
V G G S A A P P G A W P W L V R L Q L G  
430 440 450 460 470 480  
GGCAGCCTCTGTGCGGCGCGCTCCTGGTAGCGGCCTCTGGGTGCTACGGCAGCGCACT  
CCGTGCGGAGACACGCGCGCGGAGGACCATCGCCGGAGGACCCAGAGTGCCGTGCGGTGA  
G Q P L C G G V L V A A S W L T A C H  
490 500 510 520 530 540  
GCTTTGTAGCGCCCCGAATGAGCTTCTGTGGACTGTGACGCTGGCAGAGGGGTCCCGGG  
CGAAACATCCCGGGGGTTACTCGAAGACACCTGACACTGCGACCGTCTCCCCAGGGCCC  
C F V G A P N E L L W T V T L A E G S R  
550 560 570 580 590 600  
GGGAGCAAGCGGAGGAGGTGCCAGTGAACCGCATCTGCCCCACCCCAAGTTTGACCCGC  
CCCTCGTTCCGCTCCTCCACGGTCACTTGGCGTAGGACGGGGTGGGTTCAAACCTGGGCG  
G E Q A E E V P V N R I L P H P K F D P  
610 620 630 640 650 660  
GGACCTTCCACAACGACCTGGCCCTGGTGCAGCTGTGGACGCCGGTGAGCCCGGGGGGT  
CCTGGAAGGTTGTGTGACCGGGACACGTGCACACTGCGGCCACTCGGGCCCCCTTA  
R T F H N D L A L V Q L W T P V S P G G  
670 680 690 700 710 720  
CGGCGCGCCCCGTGTGCTGCCCCAGGAGCCCCAGGAGCCCCCTGCCGGAACCGCCTGCG  
GCCGCGCGGGGCACACGGACGGGGTCTCGGGGTCTCGGGGACGGCCCTTGGCGGACGC  
S A R P V C L P Q E P Q E P A G T A C  
730 740 750 760 770 780  
CCATCGCGGGGTGGGGCGCCCTCTTGAAGACGGGCTGAGGCTGAAGCAGTAGAGAGG  
GGTAGCGCCGACCCCCCGGGAGAAGCTTCTGCGCGGACTCCGACTCGTCACTCTTCC  
A I A G W G A L F E D G P E A E A V R E  
790 800 810 820 830 840  
CCCGTGTTCCTGCTCACCGACACCTGCGGAGGACCTTGGGCCCCGGGCTGCGCC  
GGGCACAAGGGGACGAGTCTGTGCTGTGGACGGCTCTCGGGACCCCGGGCCCCGACGCGG  
A R V P L L S T D T C R G A L G P G L R  
850 860 870 880 890 900  
CCAGCACCATGCTCTGCGCGGGGTACCTGGCGGGGGGCGCTTGACTCGTGCCAGGGTGACT  
GTCGTGGTACGAGACCGGCCATGGACGCCCCCCGCAACTGAGCACGCTCCCACTGA  
P S T M L C A G Y L A G G V D S C Q G D  
910 920 930 940 950 960  
CGGGAGGCCCCCTGACCTGTTCTGAGCCTGGCCCCCGCCCTAGAGAGGTCTGTTCCGAG  
GCCCTCCGGGGGACTGGACAAGACTCGGACCGGGGGCGGGATCTCTCCAGGACAAGCCTC  
S G G P L T C S E P G P R P R E V L F G  
970 980 990 1000 1010 1020  
TCACCTCCTGGGGGGACGGCTGCGGGGAGCCAGGGAAGCCCGGGGTCTACACCGCGTGG  
AGTGGAGGACCCCTGCCGACGCCCTCGGTCCCTTCGGGCCCCAGATGTGGCGCAC  
V T S W G D G C G E P G K P G V Y T R V  
1030 1040 1050 1060 1070 1080  
CAGTGTTCGAAGGACTGGCTCCAGGAGCAGATGAGCGCAGCCTCTCCAGCGCGAGCCCA  
GTCACAAGTTCTGACCGAGGTCTCTGCTACTCGCGTCGGAGGAGGTTCGGCGCTCGGT  
A V F K D W L Q E Q M S A A S S S R E P  
1090 1100 1110 1120 1130 1140  
GCTGCGAGGAGCTTCTGGCCTGGGACCCCCCAGGAGCTGCAGGCAGAGCGCGCCGGCG  
CGACGTCCCTCGAAGACCGGACCCTGGGGGGGGTCTCGACGTCCTGTCTCGGGCGGGCCG  
S C R E L E A W D P P P Q R A Q D A A P A  
1150 1160 1170 1180 1190 1200  
TCTGCGCCTTCTATGCCCCCTGTGCCCCGGGTCCAGGGCGCCTGTGCGCGCCTGGCGC  
AGACCGGGAAGATACGGGCGGACACGGGCCCCAGGGTCCCGCGGACACGCGCGGACCGCG  
L C A F Y A R L C P G S Q G A C A R L A  
1210 1220 1230 1240 1250 1260  
ACCAGCAGTGCCTGCAGCGCGCGCGGCGATGCGGTGAGTTCTGTTACCCGGACCCGGAC  
TGCTGCTACGAGCAGTTCGCGGGCGCGCTACGCCAGTCAAGACAAGTGGGCTGGGCTG  
H Q C L Q R R C R C G Q F C S P G P G  
1270 1280 1290 1300 1310 1320  
GGGGGGCAGAGGGGAGGGGCGCTGGCCAGCCTCTGACCCCGCTCCGACTCCTGTCCGGT  
CCCCCGTCTCCCCCTCCCCCGGACCGGTGCGGAGACTGGCGGGCAGGCTGAGGACAGGCCA  
R G A E G R G P G Q P L T A A P T P V R  
1330 1340 1350 1360 1370 1380  
CCGACAGAGTGCACCTCGCTGGCGCACACGCTGCTGGCCCTGCTGCGGAACCGCGCAGGAGC  
GGCGTCTCGACGTGAGCGACCGCGTGTGCGACGACCCGGACGACGCTTGGCGGTCTCTCG  
S A E L H S L A H T L S L G L L R N A Q E

-149-

1390 1400 1410 1420 1430 1440  
 TGCTCGGGCCGCGTCCGGGACTGCGGCGCCTGGCCCCCGCCCTGGCTCTCCCGCTCCAG  
 ACAGAGCCCGGCGCAGGCCCTGACGCCGCGGACCGGGGCGGGACCGAGAGGGGCGAGGTC  
 L L G P R P G L R R L A P A L A P A P  
 1450 1460 1470 1480 1490 1500  
 CGCTCAGGGAGTCTCCTCTGCACCCCGCGGAGCTGCGGCTTCACTCAGGATCGCGGG  
 GCGAGTCCCTCAGAGGAGACGTGGGGCGGGCCCTCGACGCCGAAGTGAGTCTTAGCGCCC  
 A L R E S P L H P A R E L R L H S G S R  
 1510 1520 1530 1540 1550 1560  
 CTGCAGGCACTCGGTTCCCGAAGCGGAGCGCGGAGCGCGGAGAGCCAAACGGCTGCC  
 GACGTCCGTGAGCCAAGGGCTTCGCCTCCGGCCTCGGCGCGCCTCTTCGGTTGCCGACGG  
 A A G T R F P K R R P E P R G E A N G C  
 1570 1580 1590 1600 1610 1620  
 CTGGGCTGGAGCCCCCTGCGACAGAAGTTGGCTGCCCTGCAGGGGGGCCATGCCTGGATCC  
 GACCCGACCTCGGGGACGCTGTCTTCAACCGACGGGACGTCCCCCGGGTACGGACCTAGG  
 P G L E P L R Q K L A A L Q G A H A W I  
 1630 1640 1650 1660 1670 1680  
 TGCAGGTCCCTCGGAGCACCTGGCCATGAACTTTCATGAGGTCTGGCAGATCTGGGCT  
 ACGTCCAGGGGAGCCTCGTGGACCGGTACTTGAAAGTACTCCAGGACCGTCTAGACCCGA  
 L Q V P S E H L A M N F H E V L A D L G  
 1690 1700 1710 1720 1730 1740  
 CCAAGACACTGACCGGGCTTTTCAGAGCCTGGGTGCGGGCAGGCTTGGGGGGCGGGCATG  
 GGTCTGTGACTGGCCCGAAAAGTCTCGGACCCACGCCCGTCCGAACCCCGGCCGTAC  
 S K T L T G L F R A W V R A G L G G R H  
 1750 1760 1770 1780 1790 1800  
 TGGCCTTCAGCGGCCTGGTGGGCTGGAGCCGGCCACACTGGCTCGCAGCCTCCCCCGGC  
 ACCGGAAGTCGCGGACCCCGACCTCGGCCGCTGTGACCGAGCGTCCGAGGGGGGCCG  
 V A F S G L V G L E P A T L A R S L P R  
 1810 1820 1830 1840 1850 1860  
 TGCTGGTCAGGCCCCCTGCAGGCCCTCCGCGTGGCTGCCCTGGCAGAAGGGGAGCCCCGAGG  
 ACGACCACGTCCGGGACGTCCGGAAGGCGCACCGACGGGACCGTCTTCCCTCGGGCTCC  
 L L V Q A L Q A F R V A A L A E G E  
 1870 1880 1890 1900 1910 1920  
 GACCCCTGGATGGATGTAGGGCAGGGGCCGGGCTGGAGAGGAAGGGGCACCCACTCA  
 CTGGGACCTACCTACATCCCGTCCCCGGGCCCCGACCTCTCCTTCCCCGTTGGTGGGTGAGT  
 G P W M D V G Q G P G L E R K G H H P L  
 1930 1940 1950 1960 1970 1980  
 ACCCTCAGGTACCCCCGCCAGGCAACCTGAGCCATGTCTGGGCCCCAGCCCCCTGGGG  
 TGGGAGTCCATGGGGGGCGGTCCGTTGGGACTCGGTACAGACCCGGGGGTCCGGGACCCC  
 N P Q V P P A R Q P \*  
 1990 2000 2010 2020 2030 2040  
 AGGACCTACTGCTCCAGGGGCTGAGAGGGGTTCCGGAGCATAATGACAACTGTCACTG  
 TCCTGGATGACGAGGTTCCCGACTCTCCCAAGCCCTCGTATTACTGTTGACAGTGAC  
 2050 2060 2070 2080 2090 2100  
 CCCCAGTGGCTGGGTGTGTGGGTGGGATGGGTGGGGTCTGGGGCCCCCGTGTCTT  
 GGGGTACCCGACCCACACACACCCACCCCTACCCACCCCGAGGACCCGGGGGGCACAGAA  
 2110 2120 2130 2140 2150 2160  
 CCCAGGTTTACAATCAGAGAATCACAGCTGGTTTAAATAAATGTTATTTTATAATACAGAA  
 GGGTCCAATGTTAGTCTCTTAGTGTGACCAAAATTATTTACAATAAATATTATGTGTCT  
 2170  
 AAAAAAAAAAGAAA  
 TTTTTTTTCTTT

55 CVSP17 protein sequence  
 Sequence Range: 1 to 636

60 MLLAVLLLLPLPSSWFAHGHPLYTRLPPSTLQVLSAQGTQALQAAQRSAQWAINRVAMEI  
 70 80 90 100 110 120  
 QHRSHCECRGSRPRPQALLQDPPEPGCGERRPSTANVTRAHGRIVGGSAAPPGAWFWLV  
 130 140 150 160 170 180  
 RLQLGGQPLCGGVLAASWVLTAAHCFVGAPELWTVTLAEGSRGQAEVVPVNRILPH  
 190 200 210 220 230 240  
 PKFDPRTFHNLDLVLQWTPVSPFGSARPVCLPQEPQEPAGTACAIAGWGALFEDGPEA  
 250 260 270 280 290 300  
 65 EAVREARVPLLSTDTCRGALGPGLRPLSTMLCAGYLAGGVDSQGDSSGGLTSCSEPGPRPR  
 310 320 330 340 350 360  
 EVLFGVTSWGDGCGEPGKPGVYTRVAVFKDWLQEQMSAASSSREPSCRELLAWDPQELQ  
 370 380 390 400 410 420  
 ADAARLCAFYARLCPGSGQACARLAHQQLQRRRCGQFCSPGPGRGAEGRGPGQLTAA

-150-

430 440 450 460 470 480  
 PTPVRS AELHSLAHTLLGLLRNAQELLGPRPGLRRLAPALALPAPALRESPLHPARELRL  
 490 500 510 520 530 540  
 HSGSRAAGTRFPKRRPEPRGEANGCPGLEPLRQKLAALQGAHAWILQVPSEHLAMNFHEV  
 550 560 570 580 590 600  
 LADLGSKTTLTGLFRAWVRAGLGGRHVAFSGLVGLEPATLARSLLPRLLVQALQAFRVAALA  
 610 620 630  
 EGEPEGPWMDVGQGPGLERKGGHPLNPQVPPARQP\*

10

**EXAMPLE 2****Expression of the protease domains**

Nucleic acid encoding each a full length CVSP17 and/or protease domain thereof can be cloned into a derivative of the *Pichia pastoris* vector pPIC9K (available from Invitrogen; see SEQ ID NO. 13), called pPCI9K, which is  
 15 introduced into a suitable *Pichia* host or other compatible host and used to express the encoded CVSP17 or portion thereof.

Plasmid pPIC9K features include the 5' AOX1 promoter fragment at 1-948; 5' AOX1 primer site at 855-875; alpha-factor secretion signal(s) at 949-1218; alpha-factor primer site at 1152-1172; multiple cloning site at  
 20 1192-1241; 3' AOX1 primer site at 1327-1347; 3' AOX1 transcription termination region at 1253-1586; HIS4 ORF at 4514-1980; kanamycin resistance gene at 5743-4928; 3' AOX1 fragment at 6122-6879; ColE1 origin at 7961-7288; and the ampicillin resistance gene at 8966-8106. The plasmid is derived from pPIC9K by eliminating the XhoI site in the kanamycin resistance  
 25 gene and the resulting vector is designated pPIC9KX.

Other vectors that can be used for expression of CVSP17 or portions thereof include, but are not limited to, insect and mammalian vectors as described, for example, above. The protein also can be expressed in *E. coli* in, for example, inclusion bodies in the cytoplasm or in the cytoplasm using the  
 30 strain Origami (*i.e.*, Origami B from Novagen, Madison WI) permit folding in the cytoplasm, and also can be expressed in the periplasmic space.

**EXAMPLE 3****Assays for identification of candidate compounds that modulate that activity of a serine protease****35 Assay for identifying inhibitors**

The ability of test compounds to act as inhibitors of catalytic activity of an CVSP17 can be assessed in an amidolytic assay. The inhibitor-induced

-151-

inhibition of amidolytic activity by a recombinant SP or the protease domain portions thereof, can be measured by IC50 values in such an assay.

An exemplary assay buffer is HBSA (10 mM Hepes, 150mM sodium chloride, pH 7.4, 0.1% bovine serum albumin). All reagents can be purchased from Sigma Chemical Co. (St. Louis, MO), unless otherwise indicated. Two IC50 assays at 30-minute (a 30-minute preincubation of test compound and enzyme) and at 0-minutes (no preincubation of test compound and enzyme) are conducted. For the IC50 assay at 30-minute, the following reagents are combined in appropriate wells of a Corning microtiter plate: 50 microliters of HBSA, 50 microliters of the test compound, diluted (covering a broad concentration range) in HBSA (or HBSA alone for the uninhibited velocity measurement), and 50 microliters of the SP or protease domain thereof diluted in buffer, yielding a final enzyme concentration of about 0.5-5 nM. Following a 30-minute incubation at ambient temperature, the assay is initiated by the addition of 50 microliters of a substrate for the particular SP (see, *e.g.*, table and discussion below), which was reconstituted in deionized water, and diluted in HBSA prior to the assay, yielding a final volume of 200 microliters and a final substrate concentration of 200-600  $\mu$ M.

For an IC50 assay at 0-minute, the same reagents are combined: 50 microliters of HBSA, 50 microliters of the test compound, diluted (covering the identical concentration range) in HBSA (or HBSA alone for uninhibited velocity measurement), and 50 microliters of the substrate, such as a chromogenic substrate. The assay is initiated by the addition of 50 microliters of SP. The final concentrations of all components are identical in both IC50 assays (at 30- and 0-minute incubations).

The initial velocity of the substrate hydrolysis is measured in both assays by, for example for a chromogenic substrate, the change in absorbance at a particular wavelength, using a Thermo Max<sub>2</sub> Kinetic Microplate Reader (Molecular Devices) over a 5 minute period, in which less than 5% of the added substrate was hydrolyzed. The concentration of added inhibitor, which caused a 50% decrease in the initial rate of hydrolysis was defined as the respective IC50 value in each of the two assays (30-and 0-minute).

-152-

**Another assay for identifying inhibitors**

Test compounds for inhibition of the protease activity of the protease domain are assayed in Costar 96 well tissue culture plates (Corning NY). Approximately 0.5-5 nM of the CVSP17 or protease domain thereof is mixed  
 5 with varying concentrations of inhibitor in 29.2 mM Tris, pH 8.4, 29.2 mM imidazole, 217 mM NaCl (100 mL final volume) and allowed to incubate at room temperature for 30 minutes. About 200-600  $\mu$ M substrate is added, and the reaction monitored in a SpectraMAX Plus microplate reader (Molecular Devices, Sunnyvale CA) by following the change in a parameter correlated with  
 10 hydrolysis, such as absorbance for a chromogenic substrate for 1 hour at 37° C.

**Alternative assay for screening CVSP17**

The protease domain of CVSP17 or full-length polypeptide or other catalytically active portion thereof is expressed in *Pichia pastoris*. Test  
 15 compounds are screened for modulation of the activity of the CVSP17 polypeptide or portion thereof. Approximately 1-20 nM CVSP17 is mixed in Costar 96 well tissue culture plates (Corning NY) with varying concentrations of test compounds and/or known inhibitors or agonists in 29.2 mM Tris, pH 8.4, 29.2 mM Imidazole, 217 mM NaCl (100  $\mu$ L final volume), and allowed to  
 20 incubate at room temperature for 30 minutes. 200-600  $\mu$ M s of a chromogenic substrate is added, and the reaction is monitored in a SpectraMAX Plus microplate reader (Molecular Devices, Sunnyvale CA) by measuring the change in absorbance at 405 nm for 30 minutes at 37°C.

**Identification of substrates**

25 Particular substrates for use in the assays can be identified empirically by testing substrates. The following list of substrates are exemplary of those that can be tested.

Substrate name	Structure
S 2366	pyroGlu-Pro-Arg-pNA.HCl
30 spectrozyme t-PA	CH <sub>3</sub> SO <sub>2</sub> -D-HHT-Gly-Arg-pNA.AcOH
N-p-tosyl-Gly-Pro-Arg-pNA	N-p-tosyl-Gly-Pro-Arg-pNA
Benzoyl-Val-Gly-Arg-pNA	Benzoyl-Val-Gly-Arg-pNA
Pefachrome t-PA	CH <sub>3</sub> SO <sub>2</sub> -D-HHT-Gly-Arg-pNA

-153-

	S 2765	N- $\alpha$ -Z-D-Arg-Gly-Arg-pNA.2HCl
	S 2444	pyroGlu-Gly-Arg-pNA.HCl
	S 2288	H-D-Ile-Pro-Arg-pNA.2HCl
	spectrozyme UK	Cbo-L-( $\gamma$ )Glu( $\alpha$ -t-BuO)-Gly-Arg-pNA.2AcOH
5	S 2302	H-D-Pro-Phe-Arg-pNA.2HCl
	S 2266	H-D-Val-Leu-Arg-pNA.2HCl
	S 2222	Bz-Ile-Glu(g-OR)-Gly-Arg-pNA.HCl R = H(50%) and R = CH <sub>3</sub> (50%)
	Chromozyme PK	Benzoyl-Pro-Phe-Arg-pNA
	S 2238	H-D-Phe-Pip-Arg-pNA.2HCl
10	S 2251	H-D-Val-Leu-Lys-pNA.2HCl
	Spectrozyme PI	H-D-Nle-HHT-Lys-pNA.2AcOH
		Pyr-Arg-Thr-Lys-Arg-AMC
		H-Arg-Gln-Arg-Arg-AMC
		Boc-Gln-Gly-Arg-AMC
15		Z-Arg-Arg-AMC
	Spectrozyme THE	H-D-HHT-Ala-Arg-pNA.2AcOH
	Spectrozyme fXIIa	H-D-CHT-Gly-Arg-pNA.2AcOH
		CVS 2081-6 (MeSO <sub>2</sub> -dPhe-Pro-Arg-pNA)
		Pefachrome fVIIa (CH <sub>3</sub> SO <sub>2</sub> -D-CHA-But-Arg-pNA)

20 pNA = para-nitranilide (chromogenic)  
 AMC = amino methyl coumarin (fluorescent)

If none of the above substrates are cleaved, a coupled assay can be used. Briefly, such assays test the ability of the protease to activate an enzyme, such as plasminogen and trypsinogen. To perform these assays, the single chain  
 25 protease is incubated with a zymogen, such as plasminogen or trypsinogen, in the presence of the a known substrate for plasmin or trypsin, such as a Spectrozyme substrate. If the single chain CVSP17 activates the zymogen, the activated enzyme, such as plasmin and trypsin, will degrade the substrate therefor.

#### 30 EXAMPLE 4

##### Other Assays

These assays are described with reference to MTSP1, but such assays can be readily adapted for use with CVSP17.

##### 35 Amidolytic Assay for Determining Inhibition of Serine Protease Activity of Matriptase or MTSP1

-154-

The ability of test compounds to act as inhibitors of rMAP catalytic activity was assessed by determining the inhibitor-induced inhibition of amidolytic activity by the MAP, as measured by  $IC_{50}$  values. The assay buffer was HBSA (10 mM Hepes, 150mM sodium chloride, pH 7.4, 0.1% bovine serum albumin). All reagents were from Sigma Chemical Co. (St. Louis, MO), unless  
5 otherwise indicated.

- Two  $IC_{50}$  assays (a) one at either 30-minutes or 60-minutes (a 30-minute or a 60-minute preincubation of test compound and enzyme) and (b) one at 0-minutes (no preincubation of test compound and enzyme) were conducted.
- 10 For the  $IC_{50}$  assay at either 30-minutes or 60-minutes, the following reagents were combined in appropriate wells of a Corning microtiter plate: 50 microliters of HBSA, 50 microliters of the test compound, diluted (covering a broad concentration range) in HBSA (or HBSA alone for uninhibited velocity measurement), and 50 microliters of the rMAP (Corvas International) diluted in  
15 buffer, yielding a final enzyme concentration of 250 pM as determined by active site filtration. Following either a 30-minute or a 60-minute incubation at ambient temperature, the assay was initiated by the addition of 50 microliters of the substrate S-2765 (N- $\alpha$ -Benzoyloxycarbonyl-D-arginyl-L-glycyl-L-arginine-p-nitroaniline dihydrochloride; DiaPharma Group, Inc.; Franklin, OH) to each well,  
20 yielding a final assay volume of 200 microliters and a final substrate concentration of 100  $\mu$ M (about 4-times  $K_m$ ). Before addition to the assay mixture, S-2765 was reconstituted in deionized water and diluted in HBSA. For the  $IC_{50}$  assay at 0 minutes; the same reagents were combined: 50 microliters of HBSA, 50 microliters of the test compound, diluted (covering the identical  
25 concentration range) in HBSA (or HBSA alone for uninhibited velocity measurement), and 50 microliters of the substrate S-2765. The assay was initiated by the addition of 50 microliters of rMAP. The final concentrations of all components were identical in both  $IC_{50}$  assays (at 30- or 60- and 0-minute).

The initial velocity of chromogenic substrate hydrolysis was measured in  
30 both assays by the change of absorbance at 405 nM using a Thermo Max<sup>®</sup> Kinetic Microplate Reader (Molecular Devices) over a 5 minute period, in which less than 5% of the added substrate was used. The concentration of added

-155-

inhibitor, which caused a 50% decrease in the initial rate of hydrolysis was defined as the respective  $IC_{50}$  value in each of the two assays (30- or 60-minutes and 0-minute).

*In vitro* enzyme assays for specificity determination

- 5           The ability of compounds to act as a selective inhibitor of matriptase activity was assessed by determining the concentration of test compound that inhibits the activity of matriptase by 50%, ( $IC_{50}$ ) as described in the above Example, and comparing  $IC_{50}$  value for matriptase to that determined for all or some of the following serine proteases: thrombin, recombinant tissue
- 10 plasminogen activator (rt-PA), plasmin, activated protein C, chymotrypsin and factor Xa.

- The buffer used for all assays was HBSA (10 mM HEPES, pH 7.5, 150 mM sodium chloride, 0.1% bovine serum albumin). The assay for  $IC_{50}$  determinations was conducted by combining in appropriate wells of a Corning
- 15 microtiter plate, 50 microliters of HBSA, 50 microliters of the test compound at a specified concentration (covering a broad concentration range) diluted in HBSA (or HBSA alone for  $V_0$  (uninhibited velocity) measurement), and 50 microliters of the enzyme diluted in HBSA. Following a 30 minute incubation at ambient temperature, 50 microliters of the substrate at the concentrations specified
- 20 below were added to the wells, yielding a final total volume of 200 microliters. The initial velocity of chromogenic substrate hydrolysis was measured by the change in absorbance at 405 nm using a Thermo Max® Kinetic Microplate Reader over a 5 minute period in which less than 5% of the added substrate was used. The concentration of added inhibitor which caused a 50% decrease in the initial
- 25 rate of hydrolysis was defined as the  $IC_{50}$  value.

Thrombin (fIIa) Assay

- Enzyme activity was determined using the chromogenic substrate, Pefachrome t-PA ( $CH_3SO_2$ -D-hexahydrotyrosine-glycyl-L-Arginine-p-nitroaniline, obtained from Pentapharm Ltd.). The substrate was reconstituted in deionized
- 30 water prior to use. Purified human  $\alpha$ -thrombin was obtained from Enzyme Research Laboratories, Inc. The buffer used for all assays was HBSA (10 mM HEPES, pH 7.5, 150 mM sodium chloride, 0.1% bovine serum albumin).

-156-

IC<sub>50</sub> determinations were conducted where HBSA (50  $\mu$ L),  $\alpha$ -thrombin (50  $\mu$ l) (the final enzyme concentration is 0.5 nM) and inhibitor (50  $\mu$ l) (covering a broad concentration range), were combined in appropriate wells and incubated for 30 minutes at room temperature prior to the addition of substrate

5 Pefachrome-t-PA (50  $\mu$ l) (the final substrate concentration is 250  $\mu$ M, about 5 times K<sub>m</sub>). The initial velocity of Pefachrome t-PA hydrolysis was measured by the change in absorbance at 405 nm using a Thermo Max® Kinetic Microplate Reader over a 5 minute period in which less than 5% of the added substrate was used. The concentration of added inhibitor which caused a 50% decrease in the

10 initial rate of hydrolysis was defined as the IC<sub>50</sub> value.

#### Factor Xa

Factor Xa catalytic activity was determined using the chromogenic substrate S-2765 (N-benzyloxycarbonyl-D-arginine-L-glycine-L-arginine-p-nitro-aniline), obtained from DiaPharma Group (Franklin, OH). All substrates were

15 reconstituted in deionized water prior to use. The final concentration of S-2765 was 250  $\mu$ M (about 5-times K<sub>m</sub>). Purified human Factor X was obtained from Enzyme Research Laboratories, Inc. (South Bend, IN) and Factor Xa (FXa) was activated and prepared from it as described [Bock, P.E., Craig, P.A., Olson, S.T., and Singh, P. *Arch. Biochem. Biophys.* 273:375-388 (1989)]. The enzyme was

20 diluted into HBSA prior to assay in which the final concentration was 0.25 nM. Recombinant tissue plasminogen activator (rt-PA) Assay

rt-PA catalytic activity was determined using the substrate, Pefachrome t-PA (CH<sub>3</sub>SO<sub>2</sub>-D-hexahydrotyrosine-glycyl-L-arginine-p-nitroaniline, obtained from Pentapharm Ltd.). The substrate was made up in deionized water followed by

25 dilution in HBSA prior to the assay in which the final concentration was 500 micromolar (about 3-times K<sub>m</sub>). Human rt-PA (Activase®) was obtained from Genentech Inc. The enzyme was reconstituted in deionized water and diluted into HBSA prior to the assay in which the final concentration was 1.0 nM.

-157-

**Plasmin Assay**

Plasmin catalytic activity was determined using the chromogenic substrate, S-2366 [L-pyroglutamyl-L-prolyl-L-arginine-p-nitroaniline hydrochloride], which was obtained from DiaPharma group. The substrate was made up in deionized water followed by dilution in HBSA prior to the assay in which the final concentration was 300 micromolar (about 2.5-times  $K_m$ ). Purified human plasmin was obtained from Enzyme Research Laboratories, Inc. The enzyme was diluted into HBSA prior to assay in which the final concentration was 1.0 nM.

**10                    Activated Protein C (aPC) Assay**

aPC catalytic activity was determined using the chromogenic substrate, Pefachrome PC (delta-carbobenzloxy-D-lysine-L-prolyl-L-arginine-p-nitroaniline dihydrochloride), obtained from Pentapharm Ltd.). The substrate was made up in deionized water followed by dilution in HBSA prior to the assay in which the final concentration was 400 micromolar (about 3-times  $K_m$ ). Purified human aPC was obtained from Hematologic Technologies, Inc. The enzyme was diluted into HBSA prior to assay in which the final concentration was 1.0 nM.

**Chymotrypsin Assay**

Chymotrypsin catalytic activity was determined using the chromogenic substrate, S-2586 (methoxy-succinyl-L-arginine-L-prolyl-L-tyrosyl-p-nitroanilide), which was obtained from DiaPharma Group. The substrate was made up in deionized water followed by dilution in HBSA prior to the assay in which the final concentration was 100 micromolar (about 9-times  $K_m$ ). Purified (3X-crystallized; CDI) bovine pancreatic alpha-chymotrypsin was obtained from Worthington Biochemical Corp. The enzyme was reconstituted in deionized water and diluted into HBSA prior to assay in which the final concentration was 0.5 nM.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

-158-

**WHAT IS CLAIMED IS:**

1. A substantially purified single chain or two chain polypeptide, comprising the protease domain of serine protease 17 (CVSP17) or a catalytically active portion thereof, wherein:
  - 5 a) the polypeptide also comprises at least 10 or more contiguous amino acids from residues 397-427 of SEQ ID No. 6 or comprises 10 or more contiguous amino acids encoded by a sequence of nucleotides that hybridizes under conditions of high stringency to a sequence of nucleotides that encodes residues 397-427 of SEQ ID No. 6; or
  - 10 b) the CVSP17 portion of the polypeptide consists essentially of the protease domain of the CVSP17 or a catalytically active portion thereof with the proviso that the protease domain does not include contiguous sequence Cys Arg Ser Thr Arg Ser (SEQ ID No. 18);
  - c) the polypeptide consists essentially of residues 19-332 of SEQ ID
  - 15 No. 6;
  - d) the polypeptide comprises the sequence of amino acids set forth in SEQ ID No. 6;
  - e) the polypeptide is encoded by a sequence of nucleotides that hybridizes under conditions of high stringency along at least 70% of its full
  - 20 length to a sequence of nucleotides than encodes a polypeptide of any of a)-e); and/or
  - f) the polypeptide has at least 60% sequence identity with a polypeptide of any of a)-e).
2. A purified polypeptide of claim 1, comprising a sequence of amino
- 25 acids set forth as amino acids 105-332 in SEQ ID No. 6 or a catalytically active portion thereof.
3. The polypeptide of claim 1 that is a substantially purified activated two chain CVSP17 polypeptide or a catalytically active portion thereof.
- 30 4. A substantially purified polypeptide that has at least 50%, 60%, 70%, 80%, 90% or 95% sequence identity with a polypeptide of any of claims 1-4.

-159-

5. A polypeptide of claim 1, wherein the CVSP17 portion thereof consists essentially of a protease domain or a catalytically active portion thereof.

6. A substantially purified polypeptide that has at least 50%, 60%, 70%, 80%, 90% or 95% sequence identity with the polypeptide of any of claims 1-5 and has within at least 1% of the catalytic activity on the same substrate as a polypeptide of any of claims 1-5.

7. The substantially purified polypeptide of any of claims 1-6 that is a human polypeptide.

8. A polypeptide of any of claims 1-7 that comprises:

10 (a) the sequence of amino acids set forth in SEQ ID No. 6 or a catalytically active portion thereof, or that is encoded by a sequence of nucleotides that:

(b) hybridizes under conditions of moderate or high stringency to nucleic acid complementary to an mRNA transcript present in a mammalian cell that encodes a CVSP17 encoded by (a);

(c) encodes a splice variant of (a); or

(d) comprises degenerate codons of the sequences of nucleotides of (a) or (b).

9. A polypeptide that is a mutein of the polypeptide of any of claims 1-8, wherein:

up to about 50% of the amino acids are replaced with another amino acid;

and the resulting polypeptide is a single chain or two chain polypeptide that has catalytic activity of at least 1% of the unmutated polypeptide.

10. The polypeptide of claim 9, wherein up to about 10% of the amino acids are replaced with another amino acid.

11. The polypeptide of claim 9 or claim 10, wherein the resulting polypeptide is a single chain or two chain polypeptide and has catalytic activity of at least 10% of the unmutated polypeptide.

12. The polypeptide of any of claims 9-11, wherein a free Cysteine in the protease domain is replaced with another amino acid.

-160-

13. The polypeptide of any of claims 9-12, wherein up to about 95% of the amino acids are conserved or are replaced by conservative amino acid substitutions.

14. The polypeptide of claims 12, wherein the replacing amino acid is  
5 a serine.

15. A nucleic acid molecule, comprising a sequence of nucleotides that encodes the polypeptide of any of claims 1-14.

16. The nucleic acid molecule of claim 15 that comprises a sequence of nucleotides selected from the group consisting of:

10 (a) a sequence of nucleotides set forth in SEQ ID No. 5 or a portion thereof;

(b) a sequence of nucleotides that hybridizes under high stringency along at least about 70% of its full length to the sequence of nucleotides set forth in SEQ ID No. 5 or a portion thereof;

15 (c) a sequence of nucleotides that has at least 60%, 70%, 80%, 90% or 95% sequence identity with (a); and

(d) a sequence of nucleotides comprising degenerate codon(s) of any of (a)-(c).

17. A vector comprising the nucleic acid molecule of claim 16.

20 18. The vector of claim 17 that is an expression vector.

19. The vector of claim 17 or claim 18 that is a eukaryotic vector.

20. The vector of claim 17 or claim 18 that is a prokaryotic vector.

21. The vector of any of claims 17-20 that includes a sequence of nucleotides that directs secretion of any polypeptide encoded by a sequence of  
25 nucleotides operatively linked thereto.

22. The vector of claim 17 or claim 18 that is a *Pichia* vector, a mammalian vector or an *E. coli* vector.

23. A cell, comprising the vector of any of claims 17-22.

24. The cell of claim 23 that is a prokaryotic cell.

30 25. The cell of claim 23 that is a eukaryotic cell.

26. The cell of claim 23 that is selected from among a bacterial cell, a yeast cell, a plant cell, an insect cell and an animal cell.

-161-

27. The cell of claim 26 that is a mammalian cell.

28. A recombinant non-human animal, wherein an endogenous gene that encodes a polypeptide of claim 1 has been deleted or inactivated by homologous recombination or insertional mutagenesis of the animal or an  
5 ancestor thereof.

29. A method for producing a polypeptide that contains a protease domain of a CVSP17 polypeptide, comprising:

culturing the cell of any of claims 23-27 under conditions whereby the encoded polypeptide is expressed by the cell; and  
10 recovering the expressed polypeptide.

30. The method of claim 29, wherein the polypeptide is secreted into the culture medium.

31. The method of claim 29, wherein the polypeptide is expressed in the cytoplasm of the host cell.

15 32. A method for producing a polypeptide, comprising:  
culturing of any of claims 23-27 under conditions whereby the encoded polypeptide is expressed by the cell; and  
recovering the expressed polypeptide.

33. The method of claim 32, wherein the polypeptide is expressed in  
20 inclusion bodies, and the method further comprises  
isolating the polypeptide from the inclusion bodies under conditions, whereby the polypeptide refolds into a proteolytically active form.

34. An antisense nucleic acid molecule that comprises at least 14, 16 or 30 contiguous nucleotides or modified nucleotides that are complementary to  
25 all or a portion of a contiguous sequence of nucleotides that encodes the sequence of amino acids set forth as 397-427 of SEQ ID No. 6.

35. A double-stranded RNA (dsRNA) molecule that comprises at acid molecule that comprises at least 21 contiguous nucleotides or modified nucleotides that are complementary to all or a portion of a contiguous sequence  
30 of nucleotides that encodes the sequence of amino acids set forth as 397-427 of SEQ ID No. 6.

-162-

36. A double-stranded RNA (dsRNA) molecule that comprises at acid molecule that comprises at least 21 contiguous nucleotides or modified nucleotides that are complementary to all or a portion of a contiguous sequence of nucleotides that encodes the sequence of amino set forth as SEQ ID No. 6.

5 37. A double-stranded RNA (dsRNA) molecule that comprises at least 21 contiguous nucleotides or modified nucleotides from the sequence of nucleotides encoding a polypeptide of any of claims 1-23.

38. The double-stranded dsRNA molecule of claim 36 or claim 37 that contains at least 8, 10, 12, 14, 15, 18, 21 contiguous nucleotides or modified  
10 nucleotides encoding all or a portion of amino acids 397-427 of SEQ ID No. 6.

39. A probe that comprises at least 14, 16 or 30 contiguous nucleotides or modified nucleotides that include all or a portion of a contiguous sequence of nucleotides that encodes the sequence of amino acids set forth as 397-427 of SEQ ID No. 6.

15 40. An antibody that specifically binds to the single chain form and/or two-chain form of a polypeptide of claim 1, or a fragment or derivative of the antibody containing a binding domain thereof, wherein the antibody is a polyclonal antibody or a monoclonal antibody.

41. The antibody of claim 40 that inhibits the enzymatic activity of the  
20 polypeptide.

42. An antibody that specifically binds to the leucine zipper portion of a polypeptide of claim 1, or a fragment or derivative of the antibody containing a binding domain thereof, wherein the antibody is a polyclonal antibody or a monoclonal antibody.

25 43. An antibody that specifically binds to activated two chain forms or active single chain protease domain of a CVSP17 polypeptide, wherein the antibody binds to the activated form or single chain form with at least 10-fold greater affinity than to an inactive form.

44. An antibody that specifically binds to activated two chain forms or  
30 active single chain protease domain of a CVSP17 polypeptide, wherein the antibody binds to the activated form or single chain form with at least 10-fold

-163-

greater affinity than to an inactive form, wherein the CVSP17 is a polypeptide of any of claims 1-14.

45. A conjugate, comprising:
- a) a polypeptide of claim 1, and
  - b) a targeting agent linked to the polypeptide directly or via a linker.
46. The conjugate of claim 45, wherein the targeting agent permits
- i) affinity isolation or purification of the conjugate;
  - ii) attachment of the conjugate to a surface;
  - iii) detection of the conjugate; or
  - iv) targeted delivery to a selected tissue or cell.
47. A combination, comprising:
- a) an agent or treatment that modulates the catalytic activity of the polypeptide of claim 1; and
  - b) another agent or treatment selected from anti-tumor and anti-angiogenic treatments and agents.
48. The combination of claim 47, wherein the modulator and the anti-tumor and/or anti-angiogenic agent are formulated in a single pharmaceutical composition or each is formulated in separate pharmaceutical compositions.
49. The combination of claim 47 or claim 48, wherein the modulator is an inhibitor.
50. The combination of any of claims 47-49, wherein the modulator is selected from among antibodies and antisense oligonucleotides and double-stranded RNA (dsRNA).
51. A solid support comprising two or more polypeptides of claim 1 linked thereto either directly or via a linker.
52. The support of claim 51, wherein the polypeptides comprise an array.
53. The support of claim 51, wherein the polypeptides comprise a plurality of different protease domains.
54. A solid support comprising two or more nucleic acid molecules of claim 15 or claim 16 or oligonucleotides portions thereof linked thereto either

-164-

directly or via a linker, wherein the oligonucleotides contain at least 16 nucleotides.

55. The support of claim 54, wherein the nucleic acid molecules comprise an array.

5 56. The support of claim 54 or claim 55, wherein the nucleic acid molecules comprise a plurality of molecules that encode different protease domains.

57. A method for identifying compounds that modulate the protease activity of a CVSP17 polypeptide, comprising:

10 contacting a CVSP17 polypeptide or a catalytically active portion thereof with a substrate that is proteolytically cleaved by the polypeptide, and, either simultaneously, before or after, adding a test compound or plurality thereof; measuring the amount of substrate cleaved in the presence of the test compound; and

15 selecting compounds that change the amount of substrate cleaved compared to a control, whereby compounds that modulate the activity of the polypeptide are identified.

58. A method for identifying compounds that modulate the protease activity of a CVSP17 polypeptide, comprising:

20 contacting a CVSP17 polypeptide or a catalytically active portion thereof with a substrate that is proteolytically cleaved by the polypeptide, and, either simultaneously, before or after, adding a test compound or plurality thereof; measuring the amount of substrate cleaved in the presence of the test compound; and

25 selecting compounds that change the amount of substrate cleaved compared to a control, whereby compounds that modulate the activity of the polypeptide are identified, wherein the CVSP17 polypeptide is a CVSP17 polypeptide of any of claims 1-14.

30 59. The method of claim 57 or claim 58, wherein the test compounds are small molecules, peptides, peptidomimetics, natural products, antibodies or fragments thereof that modulate the activity of the polypeptide.

-165-

60. The method of any of claims 57-59 wherein a plurality of the test compounds are screened simultaneously.

61. The method of any of claims 57-59, wherein the change in the amount of substrate cleaved is assessed by comparing the amount of substrate  
5 cleaved in the presence of the test compound with the amount of substrate cleaved in the absence of the test compound.

62. The method of claim 60 or claim 61, wherein a plurality of the polypeptides are linked to a solid support, either directly or via a linker.

63. The method of claim 62, wherein the polypeptides comprise an  
10 array.

64. A method of identifying a compound that specifically binds to a single-chain and/or two-chain protease domain and/or to single or two-chain polypeptide and/or to a proteolytically active portion of the single or two chain form thereof of a CVSP17 polypeptide, comprising:

15           contacting a CVSP17 polypeptide or a proteolytically active portion thereof with a test compound or plurality thereof under conditions conducive to binding thereof; and either:

          a) identifying test compounds that specifically bind to the single chain and two chain form of the polypeptide or to a single chain or to a two  
20 chain form thereof or to a proteolytically active portion of the single and/or two chain forms thereof, or

          b) identifying test compounds that inhibit binding of a compound known to bind a single chain and two chain form of the polypeptide or to a single or a two chain form thereof or to a proteolytically active portion of the  
25 single and/or two chain form thereof, wherein the known compound is contacted with the polypeptide before, simultaneously with or after the test compound.

65. A method of identifying a compound that specifically binds to a single-chain and/or two-chain protease domain and/or to single or two-chain polypeptide and/or to a proteolytically active portion of the single or two chain  
30 form thereof of a CVSP17 polypeptide, comprising:

-166-

contacting a CVSP17 polypeptide or a proteolytically active portion thereof with a test compound or plurality thereof under conditions conducive to binding thereof; and either:

a) identifying test compounds that specifically bind to the single  
5 chain and two chain form of the polypeptide or to a single chain or to a two chain form thereof or to a proteolytically active portion of the single and/or two chain forms thereof, or

b) identifying test compounds that inhibit binding of a compound known to bind a single chain and two chain form of the polypeptide or to a  
10 single or a two chain form thereof or to a proteolytically active portion of the single and/or two chain form thereof, wherein the known compound is contacted with the polypeptide before, simultaneously with or after the test compound, wherein the CVSP17 polypeptide is a CVSP17 polypeptide of any of claims 1-14.

15 66. The method of claim 64 or claim 65, wherein the polypeptide is linked either directly or indirectly via a linker to a solid support.

67. The method of any of claims 64-66, wherein the test compounds are small molecules, peptides, peptidomimetics, natural products, antibodies or fragments thereof.

20 68. The method of any of claims 64-67, wherein a plurality of the test substances are screened simultaneously.

69. The method of claim 68, wherein a plurality of the polypeptides are linked to a solid support.

25 70. A method for identifying activators of the zymogen form of a CVSP17, comprising:

contacting a zymogen form of a CVSP17 polypeptide or a proteolytically active portion thereof with a substrate of the activated form of the polypeptide;

30 adding a test compound, wherein the test compound is added before, after or simultaneously with the addition of the substrate; and detecting cleavage of the substrate, thereby identifying compounds that activate the zymogen.

-167-

71. A method of diagnosing the presence of a pre-malignant lesion, a malignancy, or other pathologic condition in a subject, comprising:  
obtaining a biological sample from the subject; and  
exposing the biological sample to a detectable agent that binds to a two-  
5 chain and/or single-chain form of a CVSP17 polypeptide, wherein the  
pathological condition is characterized by the presence or absence of the two-  
chain or single-chain form, wherein the CVSP17 polypeptide is a CVSP17  
polypeptide of any of claims 1-14.
72. The method of claim 70 or claim 71, wherein the substrate is a  
10 chromogenic substrate.
73. The method of any of claims 70-72, wherein the test compound is  
a small molecule, a nucleic acid or a polypeptide.
74. A method for treating or preventing a neoplastic disease, in a  
mammal, comprising administering to a mammal an effective amount of a  
15 modulator of the proteolytic activity of a polypeptide of claim 1.
75. The method of claim 74, wherein the modulator is an inhibitor.
76. The method of claim 74 or claim 75, wherein the modulator is an  
antibody that specifically binds to the polypeptide, or a fragment or derivative of  
the antibody containing a binding domain thereof, wherein the antibody is a  
20 polyclonal antibody or a monoclonal antibody.
77. A method of inhibiting tumor initiation, growth or progression or  
treating a malignant or pre-malignant condition, comprising administering an  
agent that inhibits activation of the zymogen form of a CVSP17 polypeptide or a  
potentially proteolytically active portion thereof or inhibits an activity of the  
25 activated form of CVSP17 or a potentially proteolytically active portion thereof.
78. The method of claim 77, wherein the condition is a condition of  
the breast, cervix, prostate, lung, ovary or colon.
79. The method of claim 77 or claim 78, wherein the agent is an  
antisense oligonucleotide, double-stranded RNA (dsRNA) or an antibody.
- 30 80. A method of inhibiting tumor initiation, growth or progression or  
treating a malignant or pre-malignant condition, comprising administering an  
agent that inhibits activation of the zymogen form of a CVSP17 polypeptide or a

-168-

potentially proteolytically active portion thereof or inhibits an activity of the activated form of CVSP17 or a potentially proteolytically active portion thereof. , wherein the CVSP17 polypeptide is a polypeptide of any of claims 1-14.

81. The method of any of claims 77-80, further comprising administering  
5 another treatment or agent selected from anti-tumor and anti-angiogenic treatments or agents.

82. A method of identifying a compound that binds to the single-chain or two-chain form of a CVSP17 polypeptide or to a proteolytically active portion of a single-chain or two-chain form of a CVSP17 polypeptide, comprising:  
10 contacting a test compound with both forms;  
determining to which form the compound binds; and  
if the compound binds to a form of polypeptide, further determining whether the compound has at least one of the following properties:  
(i) inhibits activation of a single-chain zymogen form of the  
15 polypeptide;  
(ii) inhibits activity of a two-chain and/or single-chain active form;  
and  
(iii) inhibits dimerization of the polypeptide.

83. A method of detecting neoplastic disease, comprising: detecting a  
20 polypeptide that comprises a polypeptide of claim 1 or a portion of a polypeptide of claim 1 in a biological sample, wherein the amount detected differs from the amount of polypeptide detected from a subject who does not have neoplastic disease.

84. The method of claim 83, wherein the biological sample is selected  
25 from the group consisting of blood, urine, saliva, tears, synovial fluid, sweat, interstitial fluid, cerebrospinal fluid, ascites fluid, tumor tissue biopsy and circulating tumor cells.

85. A method of diagnosing the presence of a pre-malignant lesion, a malignancy, or other pathologic condition in a subject, comprising:  
30 obtaining a biological sample from the subject; and  
exposing the biological sample to a detectable agent that binds to a two-chain and/or single-chain form of a CVSP17 polypeptide, wherein the

-169-

pathological condition is characterized by the presence or absence of the two-chain or single-chain form.

86. A method of diagnosing the presence of a pre-malignant lesion, a malignancy, or other pathologic condition in a subject, comprising:

5 obtaining a biological sample from the subject; and

exposing the biological sample to a detectable agent that binds to a two-chain and/or single-chain form of a CVSP17 polypeptide, wherein the pathological condition is characterized by the presence or absence of the two-chain or single-chain form, wherein the CVSP17 polypeptide is a CVSP17

10 polypeptide of any of claims 1-14.

87. A method of monitoring tumor progression and/or therapeutic effectiveness, comprising detecting and/or quantifying the level and/or activity CVSP17 polypeptide in a body tissue or fluid sample.

88. A method of monitoring tumor progression and/or therapeutic  
15 effectiveness, comprising detecting and/or quantifying the level and/or activity CVSP17 polypeptide in a body tissue or fluid sample, wherein the CVSP17 polypeptide is a CVSP17 polypeptide of any of claims 1-14.

89. The method of claim 87 or claim 88, wherein the tumor is a tumor of the breast, cervix, prostate, lung, ovary or colon.

20 90. The method of any of claims 87-89, wherein the body fluid is blood, urine, sweat, saliva, cerebrospinal fluid and synovial fluid.

91. A transgenic non-human animal, comprising heterologous nucleic acid encoding a polypeptide of claim 1.

92. A polypeptide comprising a portion of a CVSP17 polypeptide,  
25 wherein the CVSP17 portion of the polypeptide consists essentially of amino acids 1-19 of SEQ ID No. 6.

93. A nucleic acid molecule encoding a polypeptide of claim 92.

## SEQUENCE LISTING

<400> 1	tcaagagcgcg cctcggggta cc atg ggg agc gat cgg gcc cgc aag ggc gga	52
	Met Gly Ser Asp Arg Ala Arg Lys Gly Gly	
	1 5 10	
ggg ggc ccg aag gac ttc ggc gcg gga ctc aag tac aac tcc cgg cac	100	
Gly Gly Pro Lys Asp Phe Gly Ala Gly Leu Lys Tyr Asn Ser Arg His		
	15 20 25	
gag aaa gtg aat ggc ttg gag gaa ggc gtg gag ttc ctg cca gtc aac	148	
Glu Lys Val Asn Gly Leu Glu Glu Gly Val Glu Phe Leu Pro Val Asn		
	30 35 40	
aac gtc aag aag gtg gaa aag cat ggc ccg ggg cgc tgg gtg gtg ctg	196	
Asn Val Lys Lys Val Glu Lys His Gly Pro Gly Arg Trp Val Val Leu		
	45 50 55	
gca gcc gtg ctg atc ggc ctc ctc ttg gtc ttg ctg ggg atc ggc ttc	244	
Ala Ala Val Leu Ile Gly Leu Leu Leu Val Leu Leu Gly Ile Gly Phe		
	60 65 70	
ctg gtg tgg cat ttg cag tac cgg gac gtg cgt gtc cag aag gtc ttc	292	
Leu Val Trp His Leu Gln Tyr Arg Asp Val Arg Val Gln Lys Val Phe		

-2-

75	80	85	90	
aat ggc tac atg agg atc aca aat gag aat ttt gtg gat gcc tac gag				340
Asn Gly Tyr Met Arg Ile Thr Asn Glu Asn Phe Val Asp Ala Tyr Glu	95	100	105	
aac tcc aac tcc act gag ttt gta agc ctg gcc agc aag gtg aag gac				388
Asn Ser Asn Ser Thr Glu Phe Val Ser Leu Ala Ser Lys Val Lys Asp	110	115	120	
gcg ctg aag ctg ctg tac agc gga gtc cca ttc ctg ggc ccc tac cac				436
Ala Leu Lys Leu Leu Tyr Ser Gly Val Pro Phe Leu Gly Pro Tyr His	125	130	135	
aag gag tcg gct gtg acg gcc ttc agc gag ggc agc gtc atc gcc tac				484
Lys Glu Ser Ala Val Thr Ala Phe Ser Glu Gly Ser Val Ile Ala Tyr	140	145	150	
tac tgg tct gag ttc agc atc ccg cag cae ctg gtg gag gag gcc gag				532
Tyr Trp Ser Glu Phe Ser Ile Pro Gln His Leu Val Glu Glu Ala Glu	155	160	165	170
cgc gtc atg gcc gag gag cgc gta gtc atg ctg ccc ccg cgg gcg cgc				580
Arg Val Met Ala Glu Glu Arg Val Val Met Leu Pro Pro Arg Ala Arg	175	180	185	
tcc ctg aag tcc ttt gtg gtc acc tca gtg gtg gct ttc ccc acg gac				628
Ser Leu Lys Ser Phe Val Val Thr Ser Val Val Ala Phe Pro Thr Asp	190	195	200	
tcc aaa aca gta cag agg acc cag gac aac agc tgc agc ttt ggc ctg				676
Ser Lys Thr Val Gln Arg Thr Gln Asp Asn Ser Cys Ser Phe Gly Leu	205	210	215	
cac gcc cgc ggt gtg gag ctg atg cgc ttc acc acg ccc ggc ttc cct				724
His Ala Arg Gly Val Glu Leu Met Arg Phe Thr Pro Gly Phe Pro	220	225	230	
gac agc ccc tac ccc gct cat gcc cgc tgc cag tgg gcc ctg cgg ggg				772
Asp Ser Pro Tyr Pro Ala His Ala Arg Cys Gln Trp Ala Leu Arg Gly	235	240	245	250
gac gcc gac tca gtg ctg agc ctc acc ttc cgc agc ttt gac ctt gcg				820
Asp Ala Asp Ser Val Leu Ser Leu Thr Phe Arg Ser Phe Asp Leu Ala	255	260	265	
tcc tgc gac gag cgc ggc agc gac ctg gtg acg gtg tac aac acc ctg				868
Ser Cys Asp Glu Arg Gly Ser Asp Leu Val Thr Val Tyr Asn Thr Leu	270	275	280	
agc ccc atg gag ccc cac gcc ctg gtg cag ttg tgt ggc acc tac cct				916
Ser Pro Met Glu Pro His Ala Leu Val Gln Leu Cys Gly Thr Tyr Pro	285	290	295	
ccc tcc tac aac ctg acc ttc cac tcc tcc cag aac gtc ctg ctc atc				964
Pro Ser Tyr Asn Leu Thr Phe His Ser Ser Gln Asn Val Leu Leu Ile	300	305	310	
aca ctg ata acc aac act gag cgg cgg cat ccc ggc ttt gag gcc acc				1012

-3-

Thr 315	Leu	Ile	Thr	Asn	Thr 320	Glu	Arg	Arg	His	Pro 325	Gly	Phe	Glu	Ala	Thr 330	
ttc	ttc	cag	ctg	cct	agg	atg	agc	agc	tgt	gga	ggc	cgc	tta	cgt	aaa	1060
Phe	Phe	Gln	Leu	Pro 335	Arg	Met	Ser	Ser	Cys 340	Gly	Gly	Arg	Leu	Arg 345	Lys	
gcc	cag	ggg	aca	ttc	aac	agc	ccc	tac	tac	cca	ggc	cac	tac	cca	ccc	1108
Ala	Gln	Gly	Thr 350	Phe	Asn	Ser	Pro 355	Tyr	Tyr	Pro	Gly	His	Tyr 360	Pro	Pro	
aac	att	gac	tgc	aca	tgg	aac	att	gag	gtg	ccc	aac	aac	cag	cat	gtg	1156
Asn	Ile	Asp 365	Cys	Thr	Trp	Asn 370	Ile	Glu	Val	Pro	Asn	Asn 375	Gln	His	Val	
aag	gtg	agc	ttc	aaa	ttc	ttc	tac	ctg	ctg	gag	ccc	ggc	gtg	cct	gcg	1204
Lys	Val	Ser 380	Phe	Lys	Phe	Phe 385	Tyr	Leu	Leu	Glu	Pro 390	Gly	Val	Pro	Ala	
ggc	acc	tgc	ccc	aag	gac	tac	gtg	gag	atc	aat	ggg	gag	aaa	tac	tgc	1252
Gly	Thr	Cys	Pro	Lys	Asp 400	Tyr	Val	Glu	Ile	Asn 405	Gly	Glu	Lys	Tyr 410	Cys	
gga	gag	agg	tcc	cag	ttc	gtc	gtc	acc	agc	aac	agc	aac	aag	atc	aca	1300
Gly	Glu	Arg	Ser 415	Gln	Phe	Val	Val	Thr	Ser 420	Asn	Ser	Asn	Lys	Ile 425	Thr	
gtt	cgc	ttc	cac	tca	gat	cag	tcc	tac	acc	gac	acc	ggc	ttc	tta	gct	1348
Val	Arg	Phe	His 430	Ser	Asp	Gln	Ser 435	Tyr	Thr	Asp	Thr	Gly	Phe 440	Leu	Ala	
gaa	tac	ctc	tcc	tac	gac	tcc	agt	gac	cca	tgc	ccg	ggg	cag	ttc	acg	1396
Glu	Tyr	Leu 445	Ser	Tyr	Asp	Ser 450	Ser	Asp	Pro	Cys	Pro	Gly 455	Gln	Phe	Thr	
tgc	cgc	acg	ggg	cgg	tgt	atc	cgg	aag	gag	ctg	cgc	tgt	gat	ggc	tgg	1444
Cys	Arg	Thr 460	Gly	Arg	Cys 465	Ile	Arg	Lys	Glu	Leu	Arg 470	Cys	Asp	Gly	Trp	
gcc	gac	tgc	acc	gac	cac	agc	gat	gag	ctc	aac	tgc	agt	tgc	gac	gcc	1492
Ala	Asp	Cys	Thr	Asp 480	His	Ser	Asp	Glu	Leu	Asn 485	Cys	Ser	Cys	Asp 490	Ala	
ggc	cac	cag	ttc	acg	tgc	aag	aac	aag	ttc	tgc	aag	ccc	ctc	ttc	tgg	1540
Gly	His	Gln	Phe 495	Thr	Cys	Lys	Asn	Lys 500	Phe	Cys	Lys	Pro	Leu	Phe 505	Trp	
gtc	tgc	gac	agt	gtg	aac	gac	tgc	gga	gac	aac	agc	gac	gag	cag	ggg	1588
Val	Cys	Asp 510	Ser	Val	Asn	Asp	Cys 515	Gly	Asp	Asn	Ser	Asp 520	Glu	Gln	Gly	
tgc	agt	tgt	ccg	gcc	cag	acc	ttc	agg	tgt	tcc	aat	ggg	aag	tgc	ctc	1636
Cys	Ser	Cys 525	Pro	Ala	Gln	Thr 530	Phe	Arg	Cys	Ser	Asn 535	Gly	Lys	Cys	Leu	
tcg	aaa	agc	cag	cag	tgc	aat	ggg	aag	gac	gac	tgt	ggg	gac	ggg	tcc	1684
Ser	Lys	Ser 540	Gln	Gln	Cys 545	Asn	Gly	Lys	Asp	Asp	Cys 550	Gly	Asp	Gly	Ser	

-4-

gac gag gcc tcc tgc ccc aag gtg aac gtc gtc act tgt acc aaa cac	1732
Asp Glu Ala Ser Cys Pro Lys Val Asn Val Thr Cys Thr Lys His	
555 560 565 570	
acc tac cgc tgc ctc aat ggg ctc tgc ttg agc aag ggc aac cct gag	1780
Thr Tyr Arg Cys Leu Asn Gly Leu Cys Leu Ser Lys Gly Asn Pro Glu	
575 580 585	
tgt gac ggg aag gag gac tgt agc gac ggc tca gat gag aag gac tgc	1828
Cys Asp Gly Lys Glu Asp Cys Ser Asp Gly Ser Asp Glu Lys Asp Cys	
590 595 600	
gac tgt ggg ctg cgg tca ttc acg aga cag gct cgt gtt gtt ggg ggc	1876
Asp Cys Gly Leu Arg Ser Phe Thr Arg Gln Ala Arg Val Val Gly Gly	
605 610 615	
acg gat gcg gat gag ggc gag tgg ccc tgg cag gta agc ctg cat gct	1924
Thr Asp Ala Asp Glu Gly Glu Trp Pro Trp Gln Val Ser Leu His Ala	
620 625 630	
ctg ggc cag ggc cac atc tgc ggt gct tcc ctc atc tct ccc aac tgg	1972
Leu Gly Gln Gly His Ile Cys Gly Ala Ser Leu Ile Ser Pro Asn Trp	
635 640 645 650	
ctg gtc tct gcc gca cac tgc tac atc gat gac aga gga ttc agg tac	2020
Leu Val Ser Ala Ala His Cys Tyr Ile Asp Arg Gly Phe Arg Tyr	
655 660 665	
tca gac ccc acg cag tgg acg gcc ttc ctg ggc ttg cac gac cag agc	2068
Ser Asp Pro Thr Gln Trp Thr Ala Phe Leu Gly Leu His Asp Gln Ser	
670 675 680	
cag cgc agc gcc cct ggg gtg cag gag cgc agg ctc aag cgc atc atc	2116
Gln Arg Ser Ala Pro Gly Val Gln Glu Arg Arg Leu Lys Arg Ile Ile	
685 690 695	
tcc cac ccc ttc ttc aat gac ttc acc ttc gac tat gac atc gcg ctg	2164
Ser His Pro Phe Phe Asn Asp Phe Thr Phe Asp Tyr Asp Ile Ala Leu	
700 705 710	
ctg gag ctg gag aaa ccg gca gag tac agc tcc atg gtg cgg ccc atc	2212
Leu Glu Leu Glu Lys Pro Ala Glu Tyr Ser Ser Met Val Arg Pro Ile	
715 720 725 730	
tgc ctg ccg gac gcc tcc cat gtc ttc cct gcc ggc aag gcc atc tgg	2260
Cys Leu Pro Asp Ala Ser His Val Phe Pro Ala Gly Lys Ala Ile Trp	
735 740 745	
gtc acg ggc tgg gga cac acc cag tat gga ggc act ggc gcg ctg atc	2308
Val Thr Gly Trp Gly His Thr Gln Tyr Gly Gly Thr Gly Ala Leu Ile	
750 755 760	
ctg caa aag ggt gag atc cgc gtc atc aac cag acc acc tgc gag aac	2356
Leu Gln Lys Gly Glu Ile Arg Val Ile Asn Gln Thr Thr Cys Glu Asn	
765 770 775	
ctc ctg ccg cag cag atc acg ccg cgc atg atg tgc gtg ggc ttc ctc	2404
Leu Leu Pro Gln Gln Ile Thr Pro Arg Met Met Cys Val Gly Phe Leu	
780 785 790	

-5-

agc ggc ggc gtg gac tcc tgc cag ggt gat tcc ggg gga ccc ctg tcc 2452  
 Ser Gly Gly Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Ser  
 795 800 805 810

agc gtg gag gcg gat ggg cgg atc ttc cag gcc ggt gtg gtg agc tgg 2500  
 Ser Val Glu Ala Asp Gly Arg Ile Phe Gln Ala Gly Val Val Ser Trp  
 815 820 825

gga gac ggc tgc gct cag agg aac aag cca ggc gtg tac aca agg ctc 2548  
 Gly Asp Gly Cys Ala Gln Arg Asn Lys Pro Gly Val Tyr Thr Arg Leu  
 830 835 840

cct ctg ttt cgg gac tgg atc aaa gag aac act ggg gta ta ggggccgggg 2599  
 Pro Leu Phe Arg Asp Trp Ile Lys Glu Asn Thr Gly Val  
 845 850 855

ccacccaaat gtgtacacct gcggggccac ccacgtccca cccagtggtg cacgcctgca 2659  
 ggctggagac tggaccgctg actgcaccag cgccccaga acatacactg tgaactcaat 2719  
 ctccaggggt ccaaatctgc ctagaaaacc tctcgcttcc tcagcctcca aagtggagct 2779  
 gggaggtaga aggggaggac actggtggtt ctactgaccc aactgggggc aaaggtttga 2839  
 agacacagcc tccccgcca gccccaagct gggcggaggc gcgtttgtgt atatctgctt 2899  
 ccctgtctg taaggagcag cggaacgga gcttcggagc ctctcagtg aaggtggtgg 2959  
 ggctgccgga tctgggctgt ggggcccttg ggccacgctc ttgaggaagc ccaggctcgg 3019  
 aggaccctgg aaaaacagacg ggtctgagac tgaaattgtt ttaccagctc ccagggtgga 3079  
 cttcagtggt tgtatttggtg taaatgggta aaacaattta tttcttttta aaaaaaaaaa 3139  
 aaaaaaaaaa 3147

&lt;210&gt; 2

&lt;211&gt; 855

&lt;212&gt; PRT

&lt;213&gt; Homo Sapien

&lt;400&gt; 2

Met Gly Ser Asp Arg Ala Arg Lys Gly Gly Gly Gly Pro Lys Asp Phe  
 1 5 10 15  
 Gly Ala Gly Leu Lys Tyr Asn Ser Arg His Glu Lys Val Asn Gly Leu  
 20 25 30  
 Glu Glu Gly Val Glu Phe Leu Pro Val Asn Asn Val Lys Lys Val Glu  
 35 40 45  
 Lys His Gly Pro Gly Arg Trp Val Val Leu Ala Ala Val Leu Ile Gly  
 50 55 60  
 Leu Leu Leu Val Leu Leu Gly Ile Gly Phe Leu Val Trp His Leu Gln  
 65 70 75 80  
 Tyr Arg Asp Val Arg Val Gln Lys Val Phe Asn Gly Tyr Met Arg Ile  
 85 90 95  
 Thr Asn Glu Asn Phe Val Asp Ala Tyr Glu Asn Ser Asn Ser Thr Glu  
 100 105 110  
 Phe Val Ser Leu Ala Ser Lys Val Lys Asp Ala Leu Lys Leu Leu Tyr  
 115 120 125  
 Ser Gly Val Pro Phe Leu Gly Pro Tyr His Lys Glu Ser Ala Val Thr  
 130 135 140  
 Ala Phe Ser Glu Gly Ser Val Ile Ala Tyr Tyr Trp Ser Glu Phe Ser  
 145 150 155 160  
 Ile Pro Gln His Leu Val Glu Glu Ala Glu Arg Val Met Ala Glu Glu  
 165 170 175  
 Arg Val Val Met Leu Pro Pro Arg Ala Arg Ser Leu Lys Ser Phe Val  
 180 185 190  
 Val Thr Ser Val Val Ala Phe Pro Thr Asp Ser Lys Thr Val Gln Arg

			195					200					205			
Thr	Gln	Asp	Asn	Ser	Cys	Ser	Phe	Gly	Leu	His	Ala	Arg	Gly	Val	Glu	
	210					215					220					
Leu	Met	Arg	Phe	Thr	Thr	Pro	Gly	Phe	Pro	Asp	Ser	Pro	Tyr	Pro	Ala	
225					230					235					240	
His	Ala	Arg	Cys	Gln	Trp	Ala	Leu	Arg	Gly	Asp	Ala	Asp	Ser	Val	Leu	
				245					250					255		
Ser	Leu	Thr	Phe	Arg	Ser	Phe	Asp	Leu	Ala	Ser	Cys	Asp	Glu	Arg	Gly	
			260				265						270			
Ser	Asp	Leu	Val	Thr	Val	Tyr	Asn	Thr	Leu	Ser	Pro	Met	Glu	Pro	His	
		275					280					285				
Ala	Leu	Val	Gln	Leu	Cys	Gly	Thr	Tyr	Pro	Pro	Ser	Tyr	Asn	Leu	Thr	
	290					295					300					
Phe	His	Ser	Ser	Gln	Asn	Val	Leu	Leu	Ile	Thr	Leu	Ile	Thr	Asn	Thr	
305				310					315						320	
Glu	Arg	Arg	His	Pro	Gly	Phe	Glu	Ala	Thr	Phe	Phe	Gln	Leu	Pro	Arg	
				325					330					335		
Met	Ser	Ser	Cys	Gly	Gly	Arg	Leu	Arg	Lys	Ala	Gln	Gly	Thr	Phe	Asn	
			340					345					350			
Ser	Pro	Tyr	Tyr	Pro	Gly	His	Tyr	Pro	Pro	Asn	Ile	Asp	Cys	Thr	Trp	
		355					360					365				
Asn	Ile	Glu	Val	Pro	Asn	Asn	Gln	His	Val	Lys	Val	Ser	Phe	Lys	Phe	
	370					375					380					
Phe	Tyr	Leu	Leu	Glu	Pro	Gly	Val	Pro	Ala	Gly	Thr	Cys	Pro	Lys	Asp	
385				390						395					400	
Tyr	Val	Glu	Ile	Asn	Gly	Glu	Lys	Tyr	Cys	Gly	Glu	Arg	Ser	Gln	Phe	
				405					410					415		
Val	Val	Thr	Ser	Asn	Ser	Asn	Lys	Ile	Thr	Val	Arg	Phe	His	Ser	Asp	
			420					425					430			
Gln	Ser	Tyr	Thr	Asp	Thr	Gly	Phe	Leu	Ala	Glu	Tyr	Leu	Ser	Tyr	Asp	
		435					440					445				
Ser	Ser	Asp	Pro	Cys	Pro	Gly	Gln	Phe	Thr	Cys	Arg	Thr	Gly	Arg	Cys	
	450					455					460					
Ile	Arg	Lys	Glu	Leu	Arg	Cys	Asp	Gly	Trp	Ala	Asp	Cys	Thr	Asp	His	
465				470						475					480	
Ser	Asp	Glu	Leu	Asn	Cys	Ser	Cys	Asp	Ala	Gly	His	Gln	Phe	Thr	Cys	
				485					490					495		
Lys	Asn	Lys	Phe	Cys	Lys	Pro	Leu	Phe	Trp	Val	Cys	Asp	Ser	Val	Asn	
			500					505					510			
Asp	Cys	Gly	Asp	Asn	Ser	Asp	Glu	Gln	Gly	Cys	Ser	Cys	Pro	Ala	Gln	
		515					520					525				
Thr	Phe	Arg	Cys	Ser	Asn	Gly	Lys	Cys	Leu	Ser	Lys	Ser	Gln	Gln	Cys	
	530					535			</							

-7-

Thr Ala Phe Leu Gly Leu His Asp Gln Ser Gln Arg Ser Ala Pro Gly  
 675 680 685  
 Val Gln Glu Arg Arg Leu Lys Arg Ile Ile Ser His Pro Phe Phe Asn  
 690 695 700  
 Asp Phe Thr Phe Asp Tyr Asp Ile Ala Leu Leu Glu Leu Glu Lys Pro  
 705 710 715 720  
 Ala Glu Tyr Ser Ser Met Val Arg Pro Ile Cys Leu Pro Asp Ala Ser  
 725 730 735  
 His Val Phe Pro Ala Gly Lys Ala Ile Trp Val Thr Gly Trp Gly His  
 740 745 750  
 Thr Gln Tyr Gly Gly Thr Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile  
 755 760 765  
 Arg Val Ile Asn Gln Thr Thr Cys Glu Asn Leu Leu Pro Gln Gln Ile  
 770 775 780  
 Thr Pro Arg Met Met Cys Val Gly Phe Leu Ser Gly Gly Val Asp Ser  
 785 790 795 800  
 Cys Gln Gly Asp Ser Gly Gly Pro Leu Ser Ser Val Glu Ala Asp Gly  
 805 810 815  
 Arg Ile Phe Gln Ala Gly Val Val Ser Trp Gly Asp Gly Cys Ala Gln  
 820 825 830  
 Arg Asn Lys Pro Gly Val Tyr Thr Arg Leu Pro Leu Phe Arg Asp Trp  
 835 840 845  
 Ile Lys Glu Asn Thr Gly Val  
 850 855

<210> 3  
 <211> 3147  
 <212> DNA  
 <213> Homo Sapien

<220>  
 <221> CDS  
 <222> (1865) ... (2590)  
 <223> Nucleic acid sequence of protease domain of MTSP1

<400> 3  
 tcaagagcgg cctcggggta ccatggggag cgatcggggc cgcaagggcg gagggggccc 60  
 gaaggacttc ggcgcgggac tcaagtacaa ctcccggcac gagaaagtga atggcttga 120  
 ggaaggcgtg gagttcctgc cagtcaacaa cgtcaagaag gtggaaaagc atggcccggg 180  
 gcgctgggtg gtgctggcag ccgtgctgat cggcctctctc ttggtcttgc tggggatcgg 240  
 cttcctggtg tggcatttgc agtaccggga cgtgcgtgtc cagaaggtct tcaatggcta 300  
 catgaggatc acaaatgaga attttgtgga tgcctacgag aactccaact ccactgagtt 360  
 tgtaagcctg gccagcaagg tgaaggacgc gctgaagctg ctgtacagcg gagtcccatt 420  
 cctgggcccc taccacaagg agtcggctgt gacggccttc agcgagggca gcgtcatcgc 480  
 ctactactgg tctgagttca gcatcccgca gcacctggtg gaggaggccg agcgcgtcat 540  
 ggccgaggag cgcgtagtca tgetgcccc cgcggcgcg cccctgaagt cctttgtggt 600  
 cacctcagtg gtggctttcc ccacggactc caaaacagta cagaggaccc aggacaacag 660  
 ctgcagcttt ggcctgcacg cccgcggtgt ggagctgatg cgcttcacca cgcccggctt 720  
 cctgacagc cctacccccg ctcatgcccc ctgaccagtg gccctgcggg gggacgccga 780  
 ctacgtgctg agcctcacct tccgcagctt tgaccttgcg tcctgcgacg agcgcggcag 840  
 cgacctggtg acggtgtaca acacctgag ccccatggag cccacgccc tgggtgcagtt 900  
 gtgtggcacc taccctccct cctacaacct gaccttccac tcctcccaga acgtcctgct 960  
 catcacactg ataaccaaca ctgagcggcg gcatcccggc tttgaggcca ccttcttcca 1020  
 gctgcctagg atgacagct gtggaggccg cttacgtaaa gccagggga cattcaacag 1080  
 cccctactac ccaggccact acccacccaa cattgactgc acatggaaca ttgaggtgcc 1140  
 caacaaccag catgtgaagg tgagcttcaa attcttctac ctgctggagc ccggcgtgcc 1200  
 tgccggcacc tgcccgaagg actacgtgga gatcaatggg gagaaatact gcggagagag 1260  
 gtcccagttc gtcgtcacca gcaacagcaa caagatcaca gttcgttcc actcagatca 1320

-8-

```

gtcctacacc gacacogget tcttagctga atacctctcc tacgactcca gtgacccatg 1380
ccccgggagc ttcacgtgcc gcacggggcg gtgtatcccg aaggagctgc gctgtgatgg 1440
ctggggccgac tgcaccgacc acagcgatga gctcaactgc agttgcgacg ccggccacca 1500
gttcacgtgc aagaacaagt tctgcaagcc cctcttctgg gtctgcgaca gtgtgaacga 1560
ctgcccagac aacagcgacg agcaggggtg cagttgtccg gcccagacct tcaggtgttc 1620
caatgggaag tgcctctcga aaagccagca gtgcaatggg aaggacgact gtggggacgg 1680
gtccgacgag gcctctgcc ccaaggtgaa cgctcgtcact tgtaccaaac acacctaccg 1740
ctgcctcaat gggctctgct tgagcaaggg caaccctgag tgtgacggga aggaggactg 1800
tagcgacggc tcagatgaga aggactgcga ctgtgggctg cggtcattca cgagacaggc 1860
tcgt gtt gtt ggg ggc acg gat gcg gat gag ggc gag tgg ccc tgg cag 1909
  Val Val Gly Gly Thr Asp Ala Asp Glu Gly Glu Trp Pro Trp Gln
    1             5             10             15

gta agc ctg cat gct ctg ggc cag ggc cac atc tgc ggt gct tcc ctc 1957
Val Ser Leu His Ala Leu Gly Gln Gly His Ile Cys Gly Ala Ser Leu
          20             25             30

atc tct ccc aac tgg ctg gtc tct gcc gca cac tgc tac atc gat gac 2005
Ile Ser Pro Asn Trp Leu Val Ser Ala Ala His Cys Tyr Ile Asp Asp
          35             40             45

aga gga ttc agg tac tca gac ccc acg cag tgg acg gcc ttc ctg ggc 2053
Arg Gly Phe Arg Tyr Ser Asp Pro Thr Gln Trp Thr Ala Phe Leu Gly
          50             55             60

ttg cac gac cag agc cag cgc agc gcc cct ggg gtg cag gag cgc agg 2101
Leu His Asp Gln Ser Gln Arg Ser Ala Pro Gly Val Gln Glu Arg Arg
          65             70             75

ctc aag cgc atc atc tcc cac ccc ttc ttc aat gac ttc acc ttc gac 2149
Leu Lys Arg Ile Ile Ser His Pro Phe Phe Asn Asp Phe Thr Phe Asp
          80             85             90             95

tat gac atc gcg ctg ctg gag ctg gag aaa ccg gca gag tac agc tcc 2197
Tyr Asp Ile Ala Leu Leu Glu Leu Glu Lys Pro Ala Glu Tyr Ser Ser
          100             105             110

atg gtg cgg ccc atc tgc ctg ccg gac gcc tcc cat gtc ttc cct gcc 2245
Met Val Arg Pro Ile Cys Leu Pro Asp Ala Ser His Val Phe Pro Ala
          115             120             125

ggc aag gcc atc tgg gtc acg ggc tgg gga cac acc cag tat gga ggc 2293
Gly Lys Ala Ile Trp Val Thr Gly Trp Gly His Thr Gln Tyr Gly Gly
          130             135             140

act ggc gcg ctg atc ctg caa aag ggt gag atc cgc gtc atc aac cag 2341
Thr Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile Arg Val Ile Asn Gln
          145             150             155

acc acc tgc gag aac ctc ctg ccg cag cag atc acg ccg cgc atg atg 2389
Thr Thr Cys Glu Asn Leu Leu Pro Gln Gln Ile Thr Pro Arg Met Met
          160             165             170             175

tgc gtg ggc ttc ctc agc ggc ggc gtg gac tcc tgc cag ggt gat tcc 2437
Cys Val Gly Phe Leu Ser Gly Gly Val Asp Ser Cys Gln Gly Asp Ser
          180             185             190

ggg gga ccc ctg tcc agc gtg gag gcg gat ggg cgg atc ttc cag gcc 2485
Gly Gly Pro Leu Ser Ser Val Glu Ala Asp Gly Arg Ile Phe Gln Ala

```

-9-

```

195      200      205
ggg gtg gtg agc tgg gga gac ggc tgc gct cag agg aac aag cca ggc 2533
Gly Val Val Ser Trp Gly Asp Gly Cys Ala Gln Arg Asn Lys Pro Gly
      210      215      220

gtg tac aca agg ctc cct ctg ttt cgg gac tgg atc aaa gag aac act 2581
Val Tyr Thr Arg Leu Pro Leu Phe Arg Asp Trp Ile Lys Glu Asn Thr
      225      230      235

ggg gta tag gggccggggc caccctaaatg tgtacacctg cggggccacc 2630
Gly Val *
240

catcgtccac cccagtgtgc acgcctgcag gctggagact ggaccgctga ctgcaccagc 2690
gccccagaa catacactgt gaactcaatc tccagggctc caaatctgcc tagaaaacct 2750
ctcgtttcct cagctccaa agtggagctg ggaggtagaa ggggaggaca ctggtggttc 2810
tactgaccca actgggggca aaggtttgaa gacacagcct ccccgccag ccccaagctg 2870
ggccgaggcg cgtttgtgta tatctgcctc cctgtctgt aaggagcagc gggaacggag 2930
cttcggagcc tcctcagtga aggtggtggg gctgccggat ctgggctgtg gggcccttgg 2990
gccacgctct tgaggaagcc caggctcggg ggaccctgga aaacagacgg gtctgagact 3050
gaaattgttt taccagctcc cagggtggac ttcagtgtgt gtatttgtgt aaatgggtaa 3110
aacaatttat ttctttttaa aaaaaaaaaa aaaaaaa 3147

<210> 4
<211> 241
<212> PRT
<213> Homo Sapien

<400> 4
Val Val Gly Gly Thr Asp Ala Asp Glu Gly Glu Trp Pro Trp Gln Val
1 5 10 15
Ser Leu His Ala Leu Gly Gln Gly His Ile Cys Gly Ala Ser Leu Ile
20 25 30
Ser Pro Asn Trp Leu Val Ser Ala Ala His Cys Tyr Ile Asp Asp Arg
35 40 45
Gly Phe Arg Tyr Ser Asp Pro Thr Gln Trp Thr Ala Phe Leu Gly Leu
50 55 60
His Asp Gln Ser Gln Arg Ser Ala Pro Gly Val Gln Glu Arg Arg Leu
65 70 75 80
Lys Arg Ile Ile Ser His Pro Phe Phe Asn Asp Phe Thr Phe Asp Tyr
85 90 95
Asp Ile Ala Leu Leu Glu Leu Glu Lys Pro Ala Glu Tyr Ser Ser Met
100 105 110
Val Arg Pro Ile Cys Leu Pro Asp Ala Ser His Val Phe Pro Ala Gly
115 120 125
Lys Ala Ile Trp Val Thr Gly Trp Gly His Thr Gln Tyr Gly Gly Thr
130 135 140
Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile Arg Val Ile Asn Gln Thr
145 150 155 160
Thr Cys Glu Asn Leu Leu Pro Gln Gln Ile Thr Pro Arg Met Met Cys
165 170 175
Val Gly Phe Leu Ser Gly Gly Val Asp Ser Cys Gln Gly Asp Ser Gly
180 185 190
Gly Pro Leu Ser Ser Val Glu Ala Asp Gly Arg Ile Phe Gln Ala Gly
195 200 205
Val Val Ser Trp Gly Asp Gly Cys Ala Gln Arg Asn Lys Pro Gly Val
210 215 220
Tyr Thr Arg Leu Pro Leu Phe Arg Asp Trp Ile Lys Glu Asn Thr Gly

```

-10-

225  
Val

230

235

240

<210> 5  
<211> 2173  
<212> DNA  
<213> Homo Sapien

<220>  
<221> CDS  
<222> (45)...(1953)  
<223> Nucleotide sequence encoding CVSP17, including  
CVSP17 protease domain (amino acids 104-332)

<400> 5  
ctagaattca gcgcgcgtga attctagccc agctcctggt cacc atg ctg ctg got 56  
Met Leu Leu Ala  
1

gtg ctg ctg ctg cta ccc ctc cca agc tca tgg ttt gcc cac ggg cac 104  
Val Leu Leu Leu Leu Pro Leu Pro Ser Ser Trp Phe Ala His Gly His  
5 10 15 20

cca ctg tac aca cgc ctg ccc ccc agc acc ctg caa gtt ctg tcg gcc 152  
Pro Leu Tyr Thr Arg Leu Pro Pro Ser Thr Leu Gln Val Leu Ser Ala  
25 30 35

cag ggg act cag gcg ttg cag gca gcc cag agg agc gcc cag tgg gca 200  
Gln Gly Thr Gln Ala Leu Gln Ala Ala Gln Arg Ser Ala Gln Trp Ala  
40 45 50

ata aac cga gtg gcg atg gag atc cag cac aga tcg cac gag tgc cga 248  
Ile Asn Arg Val Ala Met Glu Ile Gln His Arg Ser His Glu Cys Arg  
55 60 65

gga tct ggg cgc ccc agg cct caa gct ctc ctc cag gac cca cct gag 296  
Gly Ser Gly Arg Pro Arg Pro Gln Ala Leu Leu Gln Asp Pro Pro Glu  
70 75 80

cca ggg ccg tgc ggc gag agg cgt ccg agc act gcc aat gtg acg cgg 344  
Pro Gly Pro Cys Gly Glu Arg Arg Pro Ser Thr Ala Asn Val Thr Arg  
85 90 95 100

gcc cac ggc cgc atc gtg ggg ggc agc gcg gcg ccg ccc ggg gcc tgg 392  
Ala His Gly Arg Ile Val Gly Gly Ser Ala Ala Pro Pro Gly Ala Trp  
105 110 115

ccc tgg ctg gtg agg ctg cag ctc ggc ggg cag cct ctg tgc ggc ggc 440  
Pro Trp Leu Val Arg Leu Gln Leu Gly Gly Gln Pro Leu Cys Gly Gly  
120 125 130

gtc ctg gta gcg gcc tcc tgg gtg ctc acg gca gcg cac tgc ttt gta 488  
Val Leu Val Ala Ala Ser Trp Val Leu Thr Ala Ala His Cys Phe Val  
135 140 145

ggc gcc ccg aat gag ctt ctg tgg act gtg acg ctg gca gag ggg tcc 536  
Gly Ala Pro Asn Glu Leu Leu Trp Thr Val Thr Leu Ala Glu Gly Ser

-11-

150	155	160	
cgg ggg gag caa gcg gag gag gtg cca gtg aac cgc atc ctg ccc cac Arg Gly Glu Gln Ala Glu Glu Val Pro Val Asn Arg Ile Leu Pro His 165 170 175 180			584
ccc aag ttt gac ccg cgg acc ttc cac aac gac ctg gcc ctg gtg cag Pro Lys Phe Asp Pro Arg Thr Phe His Asn Asp Leu Ala Leu Val Gln 185 190 195			632
ctg tgg acg ccg gtg agc ccg ggg gga tgg gcg cgc ccc gtg tgc ctg Leu Trp Thr Pro Val Ser Pro Gly Ser Ala Arg Pro Val Cys Leu 200 205 210			680
ccc cag gag ccc cag gag ccc cct gcc gga acc gcc tgc gcc atc gcg Pro Gln Glu Pro Gln Glu Pro Pro Ala Gly Thr Ala Cys Ala Ile Ala 215 220 225			728
ggc tgg ggc gcc ctc ttc gaa gac ggg cct gag gct gaa gca gtg aga Gly Trp Gly Ala Leu Phe Glu Asp Gly Pro Glu Ala Glu Ala Val Arg 230 235 240			776
gag gcc cgt gtt ccc ctg ctc agc acc gac acc tgc cga gga gcc ctg Glu Ala Arg Val Pro Leu Leu Ser Thr Asp Thr Cys Arg Gly Ala Leu 245 250 255 260			824
ggg ccc ggg ctg cgc ccc agc acc atg ctc tgc gcc ggg tac ctg gcg Gly Pro Gly Leu Arg Pro Ser Thr Met Leu Cys Ala Gly Tyr Leu Ala 265 270 275			872
ggg ggc gtt gac tgg tgc cag ggt gac tgg gga ggc ccc ctg acc tgt Gly Gly Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Thr Cys 280 285 290			920
tct gag cct ggc ccc cgc cct aga gag gtc ctg ttc gga gtc acc tcc Ser Glu Pro Gly Pro Arg Pro Arg Glu Val Leu Phe Gly Val Thr Ser 295 300 305			968
tgg ggg gac ggc tgc ggg gag cca ggg aag ccc ggg gtc tac acc cgc Trp Gly Asp Gly Cys Gly Glu Pro Gly Lys Pro Gly Val Tyr Thr Arg 310 315 320			1016
gtg gca gtg ttc aag gac tgg ctc cag gag cag atg agc gca gcc tcc Val Ala Val Phe Lys Asp Trp Leu Gln Glu Gln Met Ser Ala Ala Ser 325 330 335 340			1064
tcc agc cgc gag ccc agc tgc agg gag ctt ctg gcc tgg gac ccc ccc Ser Ser Arg Glu Pro Ser Cys Arg Glu Leu Leu Ala Trp Asp Pro Pro 345 350 355			1112
cag gag ctg cag gca gac gcc gcc cgg ctc tgc gcc ttc tat gcc cgc Gln Glu Leu Gln Ala Asp Ala Ala Arg Leu Cys Ala Phe Tyr Ala Arg 360 365 370			1160
ctg tgc ccg ggg tcc cag ggc gcc tgt gcg cgc ctg gcg cac cag cag Leu Cys Pro Gly Ser Gln Gly Ala Cys Ala Arg Leu Ala His Gln Gln 375 380 385			1208
tgc ctg cag cgc cgg cgg cga tgc ggt cag ttc tgt tca ccc gga ccc			1256

-12-

Cys 390	Leu	Gln	Arg	Arg	Arg	Arg 395	Cys	Gly	Gln	Phe	Cys 400	Ser	Pro	Gly	Pro	
gga 405	cgg Arg	ggg Gly	gca Ala	gag Glu	ggg 410	agg Arg	ggg Gly	cct Pro	ggc Gly	cag 415	cct Gln	ctg Pro	acc Leu	gcc Thr	gct Ala	1304 Ala
cgg Pro	act Thr	cct Pro	gtc Val	cgg 425	tcc Ser	gca Ala	gag Glu	ctg Leu	cac 430	tcg His	ctg Leu	gcg Ala	cac His	acg 435	ctg Thr	1352 Leu
ctg Leu	ggc Gly	ctg Leu	ctg Leu	cgg 440	aac Arg	gag Asn	cag Ala	gag Gln	ctg 445	ctc Leu	ggg Leu	ccg Pro	cgt Arg	ccg Pro	gga Gly	1400
ctg Leu	cgg Arg	cgc Arg	ctg Leu	gcc Ala	ccc Pro	gcc Ala	ctg Leu	gct Ala	ctc Leu	ccc Pro	gct Ala	cca Pro	gcg Ala	ctc Leu	agg Arg	1448
gag Glu	tct Ser	cct Pro	ctg Leu	cac His	ccc Pro	gcc Ala	cgg 475	gag Arg	ctg Leu	cgg 480	ctt Leu	cac His	tca Ser	gga Gly	tcg Ser	1496
cgg 485	gct Arg	gca Ala	ggc Gly	act Thr	cgg 490	ttc Arg	ccg Phe	aag Pro	cgg Lys	agg 495	ccg Arg	gag Pro	ccg Glu	cgc Pro	gga Arg	1544 Gly
gaa Glu	gcc Ala	aac Asn	ggc Gly	tgc Cys	cct 505	ggg Pro	ctg Gly	gag Leu	ccc Glu	ctg 510	cga Pro	cag Arg	aag Gln	ttg Lys	gct Leu	1592 Ala
gcc Ala	ctg Leu	cag Gln	ggg Gly	gcc Ala	cat His	gcc Ala	tgg Trp	atc Ile	ctg 525	cag Leu	gtc Gln	ccc Val	tcg Pro	gag Ser	cac Glu	1640 His
ctg Leu	gcc Ala	atg Met	aac Asn	ttt Phe	cat His	gag Glu	gtc Val	ctg 540	gca Leu	gat Ala	ctg Asp	ggc Leu	tcc Gly	aag Ser	aca Lys	1688 Thr
ctg Leu	acc Thr	ggg Gly	ctt Leu	ttc Phe	aga Arg	gcc Ala	tgg Trp	gtg Val	cgg 555	gca Arg	ggc Ala	ttg Gly	ggg Leu	ggc Gly	cgg Arg	1736
cat 565	gtg Val	gcc Ala	ttc Phe	agc Ser	ggc 570	ctg Gly	gtg Leu	ggc Val	ctg Gly	gag Leu	ccg Glu	gcc Pro	aca Ala	ctg Thr	gct Leu	1784 Ala
cgc Arg	agc Ser	ctc Leu	ccc Pro	cgg 585	ctg Arg	ctg Leu	gtg Leu	cag Val	gcc Gln	ctg 590	cag Ala	gcc Leu	ttc Gln	cgc Ala	gtg Arg	1832 Val
gct Ala	gcc Ala	ctg Leu	gca Ala	gaa Glu	ggg 600	gag Gly	ccc Glu	gag 605	gga Gly	ccc Pro	tgg Trp	atg Met	gat Asp	gta Val	ggg Gly	1880
cag Gln	ggg Gly	ccc Pro	ggg Gly	ctg Leu	gag Glu	agg Arg	aag Lys	ggg 620	cac Gly	cac His	cca Pro	ctc Leu	aac Asn	cct Pro	cag Gln	1928

-13-

gta ccc ccc gcc agg caa ccc tga g ccattgtctgg gccccagcc  
 Val Pro Pro Ala Arg Gln Pro \*  
 630 635

1973

cctggggagg acctactgct cccaggggct gagaggggtt cgggagcata atgacaaaact 2033  
 gtcactgccc cagtggctgg gtgtgtgtgg gtgggatggg gtgggggtcc tgggcccccc 2093  
 gtgtcttccc aggtttacaa tcagagaatc acagctgggt taataaatgt tatttataat 2153  
 acacagaaaa aaaaaagaaa 2173

&lt;210&gt; 6

&lt;211&gt; 635

&lt;212&gt; PRT

&lt;213&gt; Homo Sapien

&lt;220&gt;

&lt;221&gt; Protease domain

&lt;222&gt; (104)...(332)

&lt;223&gt; CVSP17

&lt;400&gt; 6

Met Leu Leu Ala Val Leu Leu Leu Leu Pro Leu Pro Ser Ser Trp Phe  
 1 5 10 15  
 Ala His Gly His Pro Leu Tyr Thr Arg Leu Pro Pro Ser Thr Leu Gln  
 20 25 30  
 Val Leu Ser Ala Gln Gly Thr Gln Ala Leu Gln Ala Gln Arg Ser  
 35 40 45  
 Ala Gln Trp Ala Ile Asn Arg Val Ala Met Glu Ile Gln His Arg Ser  
 50 55 60  
 His Glu Cys Arg Gly Ser Gly Arg Pro Arg Pro Gln Ala Leu Leu Gln  
 65 70 75 80  
 Asp Pro Pro Glu Pro Gly Pro Cys Gly Glu Arg Arg Pro Ser Thr Ala  
 85 90 95  
 Asn Val Thr Arg Ala His Gly Arg Ile Val Gly Gly Ser Ala Ala Pro  
 100 105 110  
 Pro Gly Ala Trp Pro Trp Leu Val Arg Leu Gln Leu Gly Gly Gln Pro  
 115 120 125  
 Leu Cys Gly Gly Val Leu Val Ala Ala Ser Trp Val Leu Thr Ala Ala  
 130 135 140  
 His Cys Phe Val Gly Ala Pro Asn Glu Leu Leu Trp Thr Val Thr Leu  
 145 150 155 160  
 Ala Glu Gly Ser Arg Gly Glu Gln Ala Glu Glu Val Pro Val Asn Arg  
 165 170 175  
 Ile Leu Pro His Pro Lys Phe Asp Pro Arg Thr Phe His Asn Asp Leu  
 180 185 190  
 Ala Leu Val Gln Leu Trp Thr Pro Val Ser Pro Gly Gly Ser Ala Arg  
 195 200 205  
 Pro Val Cys Leu Pro Gln Glu Pro Gln Glu Pro Pro Ala Gly Thr Ala  
 210 215 220  
 Cys Ala Ile Ala Gly Trp Gly Ala Leu Phe Glu Asp Gly Pro Glu Ala  
 225 230 235 240  
 Glu Ala Val Arg Glu Ala Arg Val Pro Leu Leu Ser Thr Asp Thr Cys  
 245 250 255  
 Arg Gly Ala Leu Gly Pro Gly Leu Arg Pro Ser Thr Met Leu Cys Ala  
 260 265 270  
 Gly Tyr Leu Ala Gly Gly Val Asp Ser Cys Gln Gly Asp Ser Gly Gly  
 275 280 285  
 Pro Leu Thr Cys Ser Glu Pro Gly Pro Arg Pro Arg Glu Val Leu Phe  
 290 295 300  
 Gly Val Thr Ser Trp Gly Asp Gly Cys Gly Glu Pro Gly Lys Pro Gly

-14-

```

y305          310          315          320
Val Tyr Thr Arg Val Ala Val Phe Lys Asp Trp Leu Gln Glu Gln Met
          325          330          335
Ser Ala Ala Ser Ser Arg Glu Pro Ser Cys Arg Glu Leu Leu Ala
          340          345          350
Trp Asp Pro Pro Gln Glu Leu Gln Ala Asp Ala Ala Arg Leu Cys Ala
          355          360          365
Phe Tyr Ala Arg Leu Cys Pro Gly Ser Gln Gly Ala Cys Ala Arg Leu
          370          375          380
Ala His Gln Gln Cys Leu Gln Arg Arg Arg Arg Cys Gly Gln Phe Cys
          385          390          395          400
Ser Pro Gly Pro Gly Arg Gly Ala Glu Gly Arg Gly Pro Gly Gln Pro
          405          410          415
Leu Thr Ala Ala Pro Thr Pro Val Arg Ser Ala Glu Leu His Ser Leu
          420          425          430
Ala His Thr Leu Leu Gly Leu Leu Arg Asn Ala Gln Glu Leu Leu Gly
          435          440          445
Pro Arg Pro Gly Leu Arg Arg Leu Ala Pro Ala Leu Ala Leu Pro Ala
          450          455          460
Pro Ala Leu Arg Glu Ser Pro Leu His Pro Ala Arg Glu Leu Arg Leu
          465          470          475          480
His Ser Gly Ser Arg Ala Ala Gly Thr Arg Phe Pro Lys Arg Arg Pro
          485          490          495
Glu Pro Arg Gly Glu Ala Asn Gly Cys Pro Gly Leu Glu Pro Leu Arg
          500          505          510
Gln Lys Leu Ala Ala Leu Gln Gly Ala His Ala Trp Ile Leu Gln Val
          515          520          525
Pro Ser Glu His Leu Ala Met Asn Phe His Glu Val Leu Ala Asp Leu
          530          535          540
Gly Ser Lys Thr Leu Thr Gly Leu Phe Arg Ala Trp Val Arg Ala Gly
          545          550          555          560
Leu Gly Gly Arg His Val Ala Phe Ser Gly Leu Val Gly Leu Glu Pro
          565          570          575
Ala Thr Leu Ala Arg Ser Leu Pro Arg Leu Leu Val Gln Ala Leu Gln
          580          585          590
Ala Phe Arg Val Ala Ala Leu Ala Glu Gly Glu Pro Glu Gly Pro Trp
          595          600          605
Met Asp Val Gly Gln Gly Pro Gly Leu Glu Arg Lys Gly His His Pro
          610          615          620
Leu Asn Pro Gln Val Pro Pro Ala Arg Gln Pro
          625          630          635

```

<210> 7  
 <211> 29  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 7  
 ctgagcctgg ccccccgcct agagaggtc

29

<210> 8  
 <211> 28  
 <212> DNA  
 <213> Artificial Sequence

-15-

<220>  
 <223> Primer  
  
 <400> 8  
 ggacaggggt cagctcaccc tctgtttg 28  
  
 <210> 9  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Primer  
  
 <400> 9  
 gagccccagg agccccctgc cggaaccgcc 30  
  
 <210> 10  
 <211> 28  
 <212> DNA  
 <213> Artificial sequence  
  
 <400> 10  
 acctctctag ggcgggggcc aggctcag 28  
  
 <210> 11  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Primer  
  
 <400> 11  
 tggcagcagt caacgcccc cgccaggtag 30  
  
 <210> 12  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Primer  
  
 <400> 12  
 tcgcgggctg gggcgccctc ttccaagacg 30  
  
 <210> 13  
 <211> 9276  
 <212> DNA  
 <213> Pichia pastoris  
  
 <400> 13  
 agatctaaca tccaaagacg aaagggttgaa tgaaaccttt ttgccatccg acatccacag 60  
 gtccattctc acacataagt gccaaaacgca acaggagggg atacactagc agcagaccgt 120  
 tgcaaacgca ggacctccac tctctttctc ctcaacaccc acttttgcca tcgaaaaacc 180  
 agcccagtta ttgggcttga ttggagctcg ctcatccaa ttcttctat taggctacta 240  
 acaccatgac tttattagcc tgtctatcct ggccccctg gcgagggttca tgtttgttta 300  
 tttccgaatg caacaagtc cgcattacac ccgaacatca ctccagatga gggctttctg 360

-16-

agtgtggggt	caaatagttt	catgttcccc	aaatggccca	aaactgacag	tttaaacgct	420
gtcttgggaac	ctaataatgac	aaaagcgtga	tctcatccaa	gatgaactaa	gtttgggtcg	480
ttgaaatgct	aacggccagt	tggtcaaaaa	gaaacttcca	aaagtcgcca	taccgtttgt	540
cttggtttggt	attgattgac	gaatgctcaa	aaataatctc	attaatgctt	agcgcagctc	600
ctctatcgct	tctgaacccc	gggtgcacotg	tgccgaaaacg	caaattgggga	aacacccgct	660
ttttggatga	ttatgcattg	tctccacatt	gtatgcttcc	aagattctgg	tggaataact	720
gctgatagcc	taacgttcat	gatcaaaatt	taactgttct	aaccctact	tgacagcaat	780
atataaacag	aagggaagctg	cctgtctta	aacctttttt	tttatcatca	ttattagctt	840
actttcataa	ttgcgactgg	ttccaattga	caagcttttg	attttaacga	cttttaacga	900
caacttgaga	agatcaaaaa	acaactaatt	attcgaaagg	tccaaacgat	gagatttcct	960
tcaattttta	ctgcagtttt	attcgcagca	tctcccgcat	tagctgctcc	agtcaacact	1020
acaacagaag	atgaaacggc	acaattcccg	gctgaagctg	tcacgcggtta	ctcagattta	1080
gaaggggatt	tcgatgttgc	tggtttgcca	ttttccaaca	gcacaaataa	cgggttattg	1140
tttataaata	ctactattgc	cagcatttgc	gctaagaag	aaggggtatc	tctcgagaaa	1200
agagaggctg	aagcttacgt	agaattccct	agggcgccg	cgaatttaatt	cgcttagac	1260
atgactgttc	ctcagttcaa	gttgggcact	tacgagaaga	ccggtcttgc	tagattctaa	1320
tcaagaggat	gtcagaatgc	catttgccctg	agagatgcag	gcttcatttt	tgatactttt	1380
ttatttgtaa	cctatatagt	ataggatttt	ttttgtcatt	ttgtttcttc	tcgtacgagc	1440
ttgctcctga	tcagcctatc	tcgcagctga	tgaatatott	gtggtagggg	tttgggaaaa	1500
tcattcgagt	ttgatgtttt	tcttggtatt	tcccactcct	cttcagagta	cagaagatta	1560
agtgagaagt	tcgtttgtgc	aagcttatcg	ataagcttta	atgcggtagt	ttatcacagt	1620
taaattgcta	acgcagtcag	gcaccgtgta	tgaaatctaa	caatgcgctc	atcgtcatcc	1680
tcggccaccgt	caccctggat	gctgtaggca	taggcttggt	tatgccggtta	ctgccgggcc	1740
tcttgccggga	tatcgctccat	tccgacagca	tcgccagtc	ctatggcggtg	ctgctagcgc	1800
tatatgcgtt	gatgcaattt	ctatgcgcac	ccgttctcgg	agcactgtcc	gaccgctttg	1860
gccgcggccc	agtcctgttc	gcttcgctac	ttggagccac	tatcgactac	gcgatcatgg	1920
cgaccacacc	cgtcctgtgg	atctatcgaa	tctaaatgta	agttaaaatc	tctaaataat	1980
taaataagtc	ccagtttctc	catacgaacc	ttaacagcat	tgccgtgagc	atctagacct	2040
tcaacagcag	ccagatccat	cactgcttgg	ccaatatggt	tcagtccctc	aggagttagc	2100
tcttgtgaag	tgatgaactt	ctggaagggt	gcagtggtta	ctccgctgta	ttgacgggca	2160
tatccgtacg	ttggc aaagt	gtgggttggt	ccggaggagt	aatctccaca	actctctgga	2220
gagtaggcac	caacaaacac	agatccagcg	tggtgtactt	gatcaacata	agaagaagca	2280
ttctcgattt	gcaggatcaa	gtgttcagga	gcgtactgat	tggaactttc	caaagcctgc	2340
tcgtagggtg	caaccgatag	ggttgttagag	tgtgcaatac	acttgcgtac	aatttcaacc	2400
cttggaactt	gcacagcttg	gttgtgaaca	gcattctcaa	ttctggcaag	ctccttgtct	2460
gtcatatcga	cagccaacag	aatcacctgg	gaatcaatac	catgttcagc	ttgagacata	2520
aggctctgagg	caacgaaatc	tggtatcagc	tatttatcag	caataactag	aacttcagaa	2580
ggcccagcag	gcattgtcaat	actacacagg	gctgatgtgt	cattttgaac	catcatcttg	2640
gcagcgttaa	cgaactgggt	tccctggacca	aatattttgt	cacacttagg	aacagtttct	2700
gttccgtaag	ccatagcagc	tactgcctgg	gcgcctcctg	ctagcacgat	acacttagca	2760
ccaaccttgt	gggcaacgta	gatgacttct	ggggtaaggg	taccatcctt	cttaggtgga	2820
gatgcaaaaa	caatttcttt	gcaaccagca	actttggcag	gaacacccag	catcagggga	2880
gtggaaggca	gaattgcggt	tccaccagga	atatagaggc	caactttctc	aataggctct	2940
gcaaaacgag	agcagactac	accagggcaa	gtctcaactt	gcaacgtctc	cgttagttag	3000
gcttcatgga	atttccctgac	gttatctata	gagagatcaa	tggtctctct	aacgttatct	3060
ggcaattgca	taagttcctc	ttggaaagga	gotttctaaca	caggtgtctt	caaagcgact	3120
ccatcaaaact	tggtcagttg	ttctaaaagg	gotttctcac	cattttgacg	aacattgtcg	3180
acaattgggt	tgactaatct	cataatctgt	tccgttttct	ggataggacg	acgaagggca	3240
tcttcaattt	cttgtgagga	ggccttagaa	acgtcaattt	tgcaaatctc	aatacgacct	3300
tcagaaggga	cttcttttagg	tttggattct	tcttttaggt	gttctctggg	gtatcctggc	3360
ttggcatctc	cttctcttct	agtgcctttt	agggacttca	tatccagggt	tctctccacc	3420
tcgtccaacg	tcacaccgta	cttggcacat	ctaactaatg	caaaataaaa	taagtcagca	3480
cattcccagg	cttatcttct	cttggattta	gcttctgcaa	gttcatcagc	ttcctcccta	3540
attttagcgt	tcaacaaaac	ttcgtcgtca	aataaccgtt	tggtataaga	accttctgga	3600
gcattgctct	tacgatccca	caagtggtgt	tccatggctc	taagaccctt	tgattggcca	3660
aaacaggaag	tgctgtccaa	gtgacagaaa	cccacacctg	tttgttcaac	cacaaatttc	3720
aagcagctct	catcacaatc	caattcgata	ccacgaact	tttgagttgc	tccagatgta	3780
gcacctttat	accacaaacc	gtgacgacga	gattggtaga	ctccagtttg	tgctcttata	3840
gcctccggaa	tagacttttt	ggacgagtag	accaggccca	acgagtaatt	agaagagtca	3900

```

gccaccaag tagtgaatag accatcgggg cggtcagtag tcaaagacgc caacaaaatt 3960
tcactgacag ggaacttttt gacatcttca gaaagtctgt attcagtagt caattgccga 4020
gcatcaataa tggggattat accagaagca acagtgggaag tcacatctac caactttgcg 4080
gtctcagaaa aagcataaac agttctacta ccgccattag tgaaactttt caaatcgccc 4140
agtgagaaag aaaaaggcac agcgatacta gcattagcgg gcaaggatgc aactttatca 4200
accaggggtcc tatagataac cctagcgccct gggatcatcc tttggacaac tctttctgcc 4260
aaatctagggt ccaaaatcac ttcatgtata ccattattgt acaacttgag caagttgtcg 4320
atcagctcct caaattgggtc ctctgtaacg gatgactcaa cttgcacatt aacttgaagc 4380
tcagtgcgatt gagtgaactt gatcagggtg tgcagctggg cagcagcata gggaaacacg 4440
gcttttccta ccaaactcaa ggaattatca aactctgcaa cacttgcgta tgcaggtagc 4500
aagggaaatg tcatacttga agtcggacag tgagtgtagt cttgagaaat tctgaagccg 4560
tatttttatt atcagttagt cagtcacacg gagatcctct acgccggacg catcgtggcc 4620
gacctgcagg gggggggggg gcgctgaggt ctgcctcgtg aagaagggtgt tgctgactca 4680
taccaggcct gaatcgcccc atcatccagc cagaaagtga gggagccacg gttgatgaga 4740
gctttgttgt aggtggacca gttggtgatt ttgaactttt gctttgccac ggaacggtct 4800
gcgttgtcgg gaagatgcgt gatctgatcc ttcaactcag caaaagtctg attttatcaa 4860
caaagccgcc gtcccgtaaa gtcagcgtaa tgctctgcca gtgttacaac caattaacca 4920
attctgatta gaaaaactca tcgagcatca aatgaaactg caatttatctc atatcaggat 4980
tatcaatacc atatttttga aaaagccgtt tctgtaatga aggagaaaaa tcaccgaggg 5040
agttccatag gatggcaaga tcttggtatc ggtctgcgat tccgactcgt ccaacatcaa 5100
tacaacctat taatttcccc tcgtcaaaaa taaggttatc aagtggaaaa tcaccatgag 5160
tgacgactga atccggtgag aatggcaaaa gcttatgcac ttctttccag acttgttcaa 5220
caggccagcc attacgctcg tcatcaaaat cactcgcatc aaccaaaccg ttattcattc 5280
gtgattgcgc ctgagcgaga cgaaatacgc gatcgctggt aaaaggacaa ttacaaacag 5340
gaatcgaatg caaccggcgc aggaacactg ccagcgcac aacaatattt tcacctgaat 5400
caggatattc ttctaatacc tggaatgctg ttttcccggg gatcgagtg gtgagtaacc 5460
atgcatcatc aggagtacgg ataaaaatgct tgatggtcgg aagaggcata aattccgtca 5520
gccagtttag tctgaccatc tcatctgtaa catcattggc aacgctacct ttgccatgtt 5580
tcagaaacaa ctctggcgca cctgggcttc catacaatcg atagattgtc gcacctgatt 5640
gcccgaacat attcgagacc catttatacc catataaatc agcatccatg ttggaattta 5700
atcgccgctc cgagcaagac gtttcccggt gaatatggct cataacaccc cttgtattac 5760
tgtttatgta agcagacagt tttattgttc atgatgatat atttttatct tgtgcaatgt 5820
aacatcagag attttgagac acaacgtggc tttccccccc cccctgcag gtcggcatca 5880
ccggcgccac aggtgcgggt gctggcgccct atatcgccga catcacgat ggggaagatc 5940
gggctcgcca cttcgggctc atgagcgctt gtttcggcgt ggggatgggt gcaggccccg 6000
tgcccggggg actgtgggc gccatctctc tgcatgcacc attccttgcc gcggcggtgc 6060
tcaacggcct caacctacta ctgggctgct tcctaattgca ggagtcgcat aaggggagagc 6120
gtcgagtatc tatgattgga gtagactcga actgtgagac cgcattcttc agtgtcttga 6180
ggtctcctat cagattatgc ccaactaaag caaccggagg aggagatttc atggtaaatt 6240
tctctgactt ttggtcatca gtagactcga actgtgagac tatctcggtt atgacagcag 6300
aaatgtcctt cttggagaca gtaaatgaag tcccaccaat aaagaaatcc ttgttatcag 6360
gaacaaaact cttgtttcga actttttcgg tgccttgaa tataaaatgt agagtggata 6420
tgtcgggtag gaatggagcg ggcaaatgct taccttctgg accttcaaga ggtatgtagg 6480
gtttgtagat actgatgcca acttcagtga caacgttgct atttcgttca aaccattccg 6540
aatccagaga aatcaaagtt gtttgtctac tattgatcca agccagtgcg gtcttgaaac 6600
tgacaatagt gtgctcgtgt tttgaggtca tctttgtatg aataaatcta gtctttgatc 6660
taaataatct tgacgagcca aggcgataaa taccctaatc taaaactctt ttaaaacggt 6720
aaaaggacaa gtatgtctgc ctgtattaaa ccccaaatca gctcgtagtc tgatcctcat 6780
caacttgagg ggcactatct tgttttagag aaatttgccg agatgcgata tcgagaaaaa 6840
ggtagcgtga ttttaaacgt gaaatttatc tcaagatctc tgccctcgcg gtttcggtga 6900
tgacgggtgaa aacctctgac acatgcagct cccggagacg gtcacagctt gtctgtaagc 6960
ggatgccggg agcagacaag cccgtcaggg cgcgtcagcg ggtgttgccg ggtgtcgggg 7020
cgcagccatg acccagtcac gtagcgatag cggagtgtat actggcttaa ctatgcggca 7080
tcagagcaga ttgtactgag agtgcccatc atgcgggtgt aaataaccga cagatgcgta 7140
aggagaaaaa accgcatcag gcgctcttcc gcttctctgc tcaactgactc gctgcgctcg 7200
gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatagc gttatccaca 7260
gaatcagggg aaagacagag tagcctaaag gccagcaaaa gggcagggaac 7320
cgtaaaaaag ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac 7380
aaaaatcgac gctcaagtca gaggtggcga aaccgcagag gactataaag ataccaggcg 7440

```

-18-

```

tttccccctg gaagetccct cgtgcgctct cctgttccga cctgcccgt taccggatac 7500
ctgtccgcct ttctcccttc gggaagcgct gcgctttctc aatgctcacg ctgtagggtat 7560
ctcagttcgg tgtagggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag 7620
cccgaccgct gcgccttctc cggtaactat cgtcttgagt ccaaccgggt aagacacgac 7680
ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt 7740
gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt 7800
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 7860
aaacaaacca ccgctggtag cgggtggtttt tttgtttgca agcagcagat tacgcgcaga 7920
aaaaaaggat ctcaagaaga tcccttgatc tttctacggg ggtctgacgc tcagtggaac 7980
gaaaaactcac gttaagggat tttgggtcatg agattatcaa aaaggatctt cacctagatc 8040
cttttaaat aaanaatgaag ttttaaatca atctaagta tatatgagta aacttggtct 8100
gacagttacc aatgcttaat cagtggagca cctatctcag cgatctgtct atttcgttca 8160
tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggagggt cttaccatct 8220
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 8280
ataaaccagc cagccggaag ggccgagcgc agaagtgggtc ctgcaacttt atccgcctcc 8340
atccagtcct ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtgtg 8400
cgcaacgttg ttgccattgc tgcaggcatc gtgggtgtcac gctcgtcgtt tgggtatggc 8460
tcattcagct ccggttccca acgatcaagg cgagttacat gatcccccat gttgtgcaaa 8520
aaagcggtta gtccttcgg toctccgac gtgtgcagaa gtaagtggc cgagtggtta 8580
tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc 8640
ttttctgtga ctgggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 8700
agttgctctt gcccgcgctc aacacgggat aataccgcgc cacatagcag aactttaaaa 8760
gtgctcatca ttggaaaacg ttcttcgggg, cgaaaaactc caaggatctt accgctgttg 8820
agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc 8880
accagcgttt ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg 8940
gcgacacgga aatggtgaat actcatactc ttcttttttc aatattattg aagcatttat 9000
cagggttatt gtctcatgag cggatacata tttgaatgta ttttagaaaa taaacaaata 9060
ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc 9120
atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtcttca agaattaatt 9180
ctcatgtttg acagcttatc atcgataagc tgactcatgt tggatttggt aaatagacgc 9240
agatcgggaa cactgaaaaa taacagttat tattcgt 9276

```

<210> 14  
 <211> 11  
 <212> PRT  
 <213> Pichia pastoris

<400> 14  
 Lys Arg Ile Ala Ser Gly Val Ile Ala Pro Lys  
 1 5 10

<210> 15  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 15  
 gccagcgtca cagtccacag aagctcatc

30

<210> 16  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence

<220>

-19-

<223> Primer

<400> 16  
ctggtcacca tgctgctggc tgtgctgctg

30

<210> 17  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 17  
gggcagcgac agtttgtcat tatgctcccg

30

<210> 18  
<211> 6  
<212> PRT  
<213>

<400> 18  
Cys Arg Ser Thr Arg Ser  
1 5